Towards Hierarchical Cluster Analysis Heatmaps as Visual Data Analysis of Entire Student Cohort Longitudinal Trajectories and Outcomes from Grade 9 through College

Alex J. Bowers
Teachers College, Columbia University
Bowers@tc.edu

Yihan Zhao
Teachers College, Columbia University
yz2663@tc.columbia.edu

Eric Ho
University of California, Los Angeles
ericmho@g.ucla.edu

ABSTRACT:
Research on data use and school Early Warning Systems (EWS) notes a central practice of researchers and practitioners is to search for patterns in student data to predict outcomes so schools can support success when students experience challenges. Yet, the domain lacks a means to visualize the rich longitudinal data that schools collect. Here, we use visual data analytic hierarchical cluster analysis (HCA) heatmaps to pattern and visualize entire longitudinal grading histories of a national sample of *n*=14,290 students from grade 9 to college in every enrolled subject and year, visualizing 6,728,920 individual datapoints. We provide both the open access code in R and an open-access online tool allowing anyone to upload their data and create a HCA heatmap, providing support for visual data analytic and data science practice for both education researchers and schooling organizations.

Keywords: cluster analysis, heatmap, early warning indicator, early warning system, data use, education data mining, education data science, visual data analytics, longitudinal data, grades, dropout, high school, post-secondary, degree, STEM

PURPOSE AND BACKGROUND:

The world outside is itself the greatest storehouse of knowledge. Human reason, drawing upon the pattern and redundancy of nature, can predict some of the consequences of human action. But the world will always remain the largest laboratory, the largest information store, from which we will learn the outcomes, good and bad, of what we have done. (Simon, 1971) (p.47)

The purpose of this study is to apply visual data analytics, specifically a hierarchical cluster analysis (HCA) heatmap, to longitudinal student trajectories of all grades and enrolled subjects from grades 9 through college as a means to inform current education early warning system research and practice through generating maps of entire cohorts of student progress linked to overall schooling outcomes. Over the last decade, there has been growing interest and research focused on education Early Warning Indicators (EWI) and Early Warning Systems (EWS) (Allensworth, 2013; Balfanz & Byrnes, 2019; Bowers, 2021b; Carl, Richardson, Cheng, Kim, & Meyer, 2013; Davis, Herzog, & Legters, 2013; Kemple, Segeritz, & Stephenson, 2013; Knowles, 2015). The goal of much of this work is to provide educators with information on the best predictors of student challenge or success early to provide an opportunity for teachers and the organization to possibly intervene and help support student success based on student needs detected from data that already exists within the system (Agasisti & Bowers, 2017; Bowers, Krumm, Feng, & Podkul, 2016; Bowers & Zhou, 2019; Farrell, 2014; Frazelle & Nagel, 2015; Krumm, Means, & Bienkowski, 2018; Piety & Pea, 2018).

Current Research on EWI/EWS that Focus on Single Variable Summary Statistics and Predictors

Throughout the EWI/EWS literature, multiple individual variables have been identified in predicting overall student outcomes, such as graduation from high school (Allensworth, Nagaoka, & Johnson, 2018; Baker, Berning, Gowda, Zhang, & Hawn, 2020; Bowers, Sprott, & Taff, 2013; Bowers & Zhou, 2019; Gubbels, van der Put, & Assink, 2019; Knowles, 2015; Rumberger, 2011). Yet, current EWI/EWS practice is to summarize student data into overall single summary statistics, such as a grade point average, overall test score, or average absences, and then to relate these summary statistics to student outcomes (Soland, 2017). This practice persists despite the findings, from across a review of 110 early warning indicators from the literature, that disaggregated clusters of student longitudinal performance trajectories are highly accurate predictors of schooling outcomes, especially in secondary school, including non-cumulative GPA (Bowers et al., 2013). While there has been much enthusiasm for informing evidence use in schools with EWI/EWS (Balfanz & Byrnes, 2019; Bowers, 2021b; Davis et al., 2013), two recent randomized controlled experiments which implemented early warning indicator monitoring across two different samples of schools in

[1] This document is a preprint of this manuscript published in the *High School Journal*. Citation:
Bowers, A.J., Zhao, Y., & Ho, E. (2022). Towards Hierarchical Cluster Analysis Heatmaps as Visual Data Analysis of Entire Student Cohort Longitudinal Trajectories and Outcomes from Grade 9 through College. *The High School Journal*, 106(1), 5-36. https://doi.org/10.1353/hsj.2022.a906700
[2] This material is based upon work supported by the National Science Foundation under Grant No. NSF IIS-1546653. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
[3] *Acknowledgements:* We thank Lucretia Glover for her contributions on this project at the beginning stages.

the USA failed to find a significant effect of this practice (Faria et al., 2017; Mac Iver, Stein, Davis, Balfanz, & Fox, 2019). Additionally, recent research from Oregon in which 65 districts implemented a grade 9 early warning system that provided monitoring on chronic absenteeism, discipline, course progression, and state test score performance, found no relationship between the adoption of the EWS and student discipline infraction rates, credit accrual, or math or English state standardized test performance (Sepanik, Zhu, Shih, & Commins, 2021).

While these results are early, in limited samples, and in need of replication, our supposition is that these findings suggest that current data summary practices which focus on bar graphs, line plots, and summary averages across schools (Bowers, 2021a; Bowers et al., 2016; Bowers, Shoho, & Barnett, 2014; Schwendimann et al., 2017; Selwyn, Pangrazio, & Cumbo, 2021) could be augmented and complemented through adding additional tools to the education data analysts' toolbox that provide additional views into the life course of students throughout their time in school (Alexander, Entwisle, & Kabbani, 2001; Bowers & Krumm, 2021; Nitkin, Ready, & Bowers, 2022; Pallas, 1993), which could ultimately help support richer and more in-depth data discussions between educators focusing on supporting individual student journeys through the system provided to the student.

As recently noted by Selwyn et al. (2021), educators in their three case study schools reported their high aspirations for data use, saying "But there's this pattern here… That sort of stuff is gold especially for a year level coordinator. If they can see that then they can start to do something." (p.78) In practice however, the authors noted that current practice relies not on finding these interesting patterns, but rather on education data analysts creating bespoke, ad hoc data summaries, often in Microsoft Excel, in which the analytics are "simple frequency counts… and modest cross-tabulations" (Selwyn et al., 2021) (p.84). Overall summary statistics aggregate students' interesting, variable, and individual journeys through the schooling system into single summary numbers, obfuscating the lived experiences of each student across their individual grades and enrollments through subjects and grade levels throughout their time in schools.

A Bottom-Up Descriptive Framework for Visualizing "Maps" of the Curricular Ecosystem

In contrast to this practice, some researchers have proposed an alternative holistic "bottom up" framework. This framework considers both the student and the school system as a dynamic curricular ecosystem that provides a rich set of information to describe the lived experiences of individual students, information which is rarely used to view and work to understand each student as an individual (Frank et al., 2008; Heck, Price, & Thomas, 2004). This "sociocurricular" system:

> …can be conceptualized as emergent structures resulting from a series of student encounters with courses… One unique feature of this type of analysis is that it allows the structure of a complex array of student, teacher, and course-taking events over time to emerge from their simultaneous analysis, with minimal assumptions required about what this structure might look like ahead of time. Importantly, it identifies actual patterns by considering each student's complete course-taking pattern… (Heck et al., 2004, pp. 327-328)

Currently there are few options for EWS practitioners to plot, view, and examine data that can be thousands or tens of thousands of rows (students) by hundreds or thousands of columns (variables collected over time). Indeed, current practice does not allow anyone working in the educational system to see the many different ways that students move through the system. This reality precludes researchers and practitioners from identifying patterns throughout these individual histories of each student's experiences with the schooling provided to them through which more efficient routes or unexpected and beneficial practices could be identified (Bowers, 2010; Bowers & Krumm, 2021). As noted in human-system interaction design theory (Preece et al., 1994), when the most efficient route through a system is not well known or studied, then providing map-like knowledge is important to generate so that the most efficient yet currently unidentified pathways to objectives can be found, especially from rarely occurring cases and at the boundaries of the system (Verplank, 2003).

Hierarchical Cluster Analysis (HCA) Heatmaps as Visual Data Analytics

In the present study, we address these issues through applying recent innovations in visual data analytics in education research and methods (Bienkowski, Feng, & Means, 2012) which draw on innovations from the fields of data science, big data, and information system design (Agasisti & Bowers, 2017; Bowers, 2021b, in press; Bowers, Bang, Pan, & Graves, 2019; Bowers & Krumm, 2021; Fischer et al., 2020; Krumm & Bowers, 2022; Krumm et al., 2018; Piety, 2019; Piety, Hickey, & Bishop, 2014; Piety & Pea, 2018; Stahl, Gabrys, Gaber, & Berendsen, 2013). Visual data analytics is "designed to help expose patterns, trends, and exceptions in very large heterogeneous and dynamic datasets collected from complex systems" (Bienkowski et al., 2012, p. 15). This makes it ideal to apply to the large system of data that schools currently collect yet rarely leverage, namely teacher-assigned grades in each subject across multiple grade levels (Bowers, 2009, 2011, 2019; Brookhart et al., 2016). Here, we apply the visual data analytic technique of hierarchical cluster analysis heatmaps to longitudinal grades data.

Used extensively in big data fields such as bioinformatics, Hierarchical Cluster Analysis (HCA) heatmaps are a visual data analytic technique that takes as input large heterogeneous datasets of many different variables or features (columns in the dataset) across many sampled participants or respondents (rows). The rows by columns data matrix is clustered such that participant rows of data that are most similar by a selected metric across the hyperdimensional dataspace are placed proximal to one another in the ordering of the rows (Romesburg, 1984; Wilkinson & Friendly, 2009). The data in each cell for each row by column (person by variable) is then visualized with a "heatmap" in which lower levels of a variable are represented by one color (here a "colder" blue), and higher levels by another color (here a "hotter" red) (Bowers, 2010; Eisen, Spellman, Brown, & Botstein, 1998). For example, over the last 20 years in big data bioinformatics cancer research and molecular biology, scientists have worked to analyze the flood of data from across medicine, such as from the sequencing of the human genome, using cluster analysis heatmaps (Gu, Eils, & Schlesner, 2016). To organize, pattern, and analyze these large sets of data, studies apply these ideas of map-like visualizations through the use of hierarchical cluster analysis heatmaps, such as in cancer research in which hundreds or thousands of gene expression levels are patterned and visualized across hundreds or thousands of human cancer patient samples, representing varying levels of disease invasiveness and varying degrees of patient prognosis, identifying patterns of specific genes related directly to better or worse outcomes (Bowers, Stanton, & Boylan, 2000; Kim, Watkinson, Varadan, & Anastassiou, 2010). In virology, for example, hierarchical cluster analysis heatmaps have been used to pinpoint patterns in the COVID 19 pandemic, examining outcomes versus patient responses to the virus (Lucas et al., 2020). In each of these domains, the cluster analysis heatmap is used to describe the patterns in the data.

Indeed, in bioinformatics, the popularity of cluster analysis heatmaps is well-known, as Weinstein (2008) noted in *Science* more than a decade ago:

> For visualization, by far the most popular graphical representation has been the clustered heat map, which compacts large amounts of information into a small space to bring out coherent patterns in the data.... Since their debut over 10 years ago, clustered heat maps have appeared in well over 4000 biological or biomedical publications (Weinstein, 2008, p. 1772).

Continuing, Weinstein (2008) provides the following useful definition of the cluster analysis heatmap, in which:

> In the case of gene expression data, the color assigned to a point in the heat map grid indicates how much of a particular RNA or protein is expressed in a given sample. The gene expression level is generally indicated by red for high expression and either green or blue for low expression. Coherent patterns (patches) of color are generated by hierarchical clustering on both horizontal and vertical axes to bring like together with like. Cluster relationships are indicated by tree-like structures adjacent to the heat map, and the patches of color may indicate functional relationships among genes and samples source of order other than clustering (for example, time in a series of measurements) (Weinstein, 2008, p.1772).

In education, HCA heatmaps have been used to pattern and visualize data ranging from personalized learning logfiles (Bowers & Krumm, 2021; Bowers et al., 2016; Krumm et al., 2018; Nitkin et al., 2022); K-12 student benchmark test performance across state standards (Adams et al., 2021); higher education online learning management system (LMS) course pageviews (Lee, Recker, Bowers, & Yuan, 2016) and student course evaluation surveys (Reverter, Martinez, Currey, van Bommel, & Hudson, 2020); the use of student virtual manipulatives in mathematics (Moyer-Packenham, Tucker, Westenskow, & Symanzik, 2015); comparing principal survey responses about their perceptions of school leadership across national contexts (Ólafsson & Hansen, 2022); and participant interaction in circuit and logic gate educational design games (Jorion et al., 2020). Importantly, in a EWI/EWS students are empirically ordered rather placed in a list by student name or ID,. Students with the most similar data patterns are next to one another in the clustered list, and when combined with the heatmap representing their data, the human eye quickly identifies clusters of similar and dissimilar data patterns – visual data analytics. Thus, while a line or bar chart of thousands of student rows by hundreds of columns of data would be uninterpretable, the same data patterned and displayed in a HCA heatmap provides a means to visualize the individual and complex patterns across the entire dataset, displaying each datapoint for each person, patterned such that overall trends as well as exception cases are pinpointed for interpretation and possible action (Bowers, 2010).

Exploratory Visual Data Analysis, Data Science, and Learning from Data

A central issue in a discussion of HCA heatmaps for any field (education, bioinformatics, or otherwise) is that compared to inferential statistics which includes analysis and model fit metrics, p-value hypothesis significance tests, and theory testing (Bland & Altman, 1986), cluster analysis heatmaps complement, yet are quite different from, these classic psychology and econometrics statistics as heatmaps are a descriptive exploratory data visualization method designed to visualize high dimensionality relationships across hundreds to thousands of rows and columns of data. Exploratory data visualization has a long history over the last half century of research methods literature, as noted across the Exploratory Data Analysis (EDA) literature (Behrens, DiCerbo, Yel, & Levy, 2012; Tukey, 1962, 1977), in which EDA functions to "address the broad question of 'what is going on here?' [with] an emphasis on graphic representations of data [. . .] [in which] the goal of EDA is to discover patterns in data" (p. 132) (Behrens, 1997). Indeed, in research on the recent impact of the important advances in statistics and data analysis over the last half century, Gelman and Vehtari (2020) note:

> Following Tukey (1962), the proponents of exploratory data analysis have emphasized the limitations of asymptotic theory and the corresponding benefits of open-ended exploration and communication (Cleveland, 1985) along with a general view of data science as going beyond statistical theory (Chambers, 1993; Donoho, 2017). This fits into a view of statistical modeling that is focused more on discovery than on the testing of fixed hypotheses, and as such has been influential not just in the development of specific graphical methods but also in moving the field of statistics away from theorem-proving and toward a more open and, we would say, healthier perspective on the role of learning from data in science. An example in medical statistics is the much-cited article by Bland and Altman (1986) that recommended graphical methods for data

> comparison in place of correlations and regressions (p.2090) (Gelman & Vehtari, 2021).

This emphasis on graphical exploratory data visualization as a useful and practical method to surface unknown patterns in the data was recently exemplified in education research in which the researchers partnered with education leaders in a school to pattern and visualize multiple strands of summative assessments for 476 students in Algebra I using HCA heatmaps visualizing more than 4,000 individual data points, in which on reflecting on the collaborative work with the data scientists a school leader noted:

> I think that this [HCA heatmap discussion] just opened up a huge frame of conversation for us to have with course-level teams and provide them data that I think they'll be able to dig deep on and start to revise a lot of these courses… It was a totally different way of visualizing data that I think we haven't seen before... It was just a really, really interesting way to think about data. Because we think about it in simpler terms here and so it's nice to see the larger possibilities with what we can do with the data that we have (p.641) (Bowers & Krumm, 2021).

Thus, we focus on exploratory data analysis as a recommended yet under-utilized descriptive research method as we bring together these ideas of description, visual data analytics, exploratory data analysis, data science, and the "from the ground up" theory of the sociocurricular system as applied to visualizing and describing curricular "maps" of student enrollment and performance over time connected to overall schooling outcomes.

HCA Heatmaps to Visualize Entire Longitudinal Grading Histories of Large Student Datasets

Specifically for the present study and EWI/EWS, our goal was to replicate and extend Bowers (2010) who applied HCA heatmaps to visualize the entire longitudinal grading history of every grade in every subject for every student K-12 from two small school districts (Bowers, 2010). Analyzing a dataset including n=188 students and each of their grades in each subject K-12 using HCA heatmaps, the author found two main clusters that each represented about 50% of the dataset, about half of the students with high grades who took the ACT, and half who had overall low grades throughout K-12 and accounted for almost all of the dropouts. For specific subclusters, Bowers (2010) identified that persistent longitudinal grade patterns could be identified as early as grade 4 that correlated with dropping out before the end of grade 12.

While currently one of the only applications of HCA heatmaps to longitudinal grades data of this type, the Bowers (2010) study is limited. First, the dataset is small and context-dependent. Second, the analysis did not include information beyond grade 12 outcomes. Third, the analysis was not performed in an open coding environment in which the code for the analytics and visualization could be shared.

Thus, in the present study we aim to replicate and extend Bowers (2010) by applying visual data analytic HCA heatmaps to fine-grained longitudinal grade and subject data for a large national sample of over ten thousand students from grades 9 through college, examining patterns linked to college and career outcomes, with an added focus on STEM outcomes (Science, Technology, Engineering, and Mathematics). To help inform EWI/EWS research and practice, we provide the full code for the analysis in the open source R software (R Development Core Team, 2019), and we also provide an open source online application written with the Shiny package in R, such that a user can download the code and run it locally in an encrypted environment, or upload a .csv or Microsoft Excel file of their data, and generate their own HCA heatmap using the visual data analytic recommendations from the literature, without the need to code (providing a low-code/no-code environment (Lethbridge, 2021)). Our research questions for this study were:

1) To what extent are student grading patterns from grade 9 through college identified through hierarchical cluster analysis heatmaps.
2) To what extent do patterns identified through HCA heatmaps link to overall schooling outcomes, such as dropping out, graduating, attending college, majoring in a STEM subject, graduating with a STEM degree, or obtaining a job in a STEM field by age 26.

METHODS:

Sample:

This study is a secondary data analysis of the restricted access Education Longitudinal Study of 2002 (ELS:2002) High School Transcript Study (HSTS), Post-Secondary Transcripts and first through third follow-ups. ELS:2002 is a survey of about 15,000 USA students who were in grade 10 in 2002, collected by the U.S. Department of Education, National Center for Education Statistics (NCES), in which the NCES collected the entire high school and college transcripts of participants as well as schooling outcomes by age 26 (Ingels et al., 2014), representing the most complete national public dataset for students through age 26 at the time of writing. Transcripts included teacher-assigned grades in each subject from grades 9 through 12, as well as post-secondary school starting in 2006 (for dual-enrollment students) through four years of college. All courses were classified by NCES into 54 standardized subjects in high school (Classification of Secondary School Courses, CSSC) and 47 subjects in college (College Course Map, CCM), representing all courses taken in high school and college. Thus, our final data matrix for analysis included n=14,290 student rows by 451 columns. Grades in each column were standardized to a 5-point grading scale (0 to 4.0) and z-scored to prevent overweighting in the data matrix.

Analysis and Visualization:

To analyze and plot the HCA heatmap, we used the ComplexHeatmap() package (Gu et al., 2016) in the open source R statistical software (R Development Core Team, 2019). We followed past recommendations for HCA heatmaps with this type of data (Bowers, 2010) using uncentered correlation as the distance metric and average linkage as the unsupervised hierarchical clustering algorithm, as it is robust to missing data issues that are inherent in this type of data as many students may not enroll in a range of subjects. This is in comparison to other clustering algorithms, such as k-means, which are much less robust to even small amounts of missing data. As noted throughout the clustering analysis research literature, cluster analysis is a descriptive method and, for hierarchical cluster analysis here, clusters are defined through the empirical hierarchical structure identified through the algorithm, as there is the long-standing issue across the cluster analysis domain of a lack of agreement on fit metrics given that clustering algorithms are designed to group similar objects rather than identify and learn separation and borders of different classes (Färber et al., 2010; Zimmermann, 2020).

Briefly, in hierarchical cluster analysis, a distance metric is used to calculate a distance matrix for each row in the dataset, then each row starts off as its own cluster, then clusters agglomerate in a hierarchical fashion, with the most similar data vectors (rows) next to each other in the reordered list, as defined by the clustering algorithm (Romesburg, 1984; for reviews of clustering algorithms see (Iam-On & Boongoen, 2017; D. Xu & Tian, 2015; R. Xu & Wunsch, 2005; Zhao & Karypis, 2005). Here, following the previous research (Bowers, 2010) we used uncentered correlation (Equation 1) as the distance metric, $r(x_i,y_i)$, which is preferred over the similar Pearson correlation as it assumes the mean is zero for each vector x and y. Thus where the Pearson correlation would have a high correlation of 1 for two data vectors that had the same shape but are offset by a constant value, the uncentered correlation will not be 1 (Anderberg, 1973). An additional useful property of the uncentered correlation is that it is equivalent to the cosine angle between two vectors. Thus, the uncentered correlation for any two vectors x_i and y_i of sample size n is:

$$r(x_i, y_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\sigma_x^{(0)}}\right)\left(\frac{y_i}{\sigma_y^{(0)}}\right) \qquad \text{Equation 1}$$

where:

$$\sigma_x^{(0)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2}$$

$$\sigma_y^{(0)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i)^2}$$

For the agglomeration clustering algorithm we then used average linkage, which iteratively starts with each case as a cluster, then, using the distances calculated above, iterates over the distance matrix, agglomerating cases and clusters into larger clusters by calculating the average distance of the total number of cases within each of two clusters, in which the two cases with the smallest average distance are joined first, the distance matrix is updated, and then the process iterates hierarchically clustering similar cases and clusters together.

We then built an RShiny application to allow upload of data files for HCA heatmap analysis and visualization (https://ohrice.shinyapps.io/Heatmap). We provide the full R code for all analysis, visualizations, and the full R Shiny application code in the online Supplement 1 Appendix: https://doi.org/10.7916/cqvn-9t71. Additionally, as we realize that HCA heatmaps may be unfamiliar to education researchers and education data analysts, and our dataset here for the results uses a restricted access dataset from NCES, we provide in the online Supplement 2 Appendix (https://doi.org/10.7916/r1mg-yn37) a specific walkthrough in R markdown using the “mtcars” dataset which is included by default in R. This markdown demonstrates each step of the HCA heatmap process with the public “toy” dataset of attributes of automobiles from the 1974 Motor Trend data (such as horsepower, miles per gallon, weight, etc.) and includes all of the code needed along with the example output generated from the code to provide an opportunity for data analysts and researchers to practice and replicate the analysis and then be able to implement this technique themselves with their own data in their own organization. Importantly for education data analysts using restricted datasets, the code can be downloaded and run locally on an individual’s computer, either encrypted or not, as well as modified and adapted by the user.

Figure 1 provides an overview of data processing and analysis, and Appendix A provides the order of subjects in the HCA heatmap by grade level. As it may be the first time that a reader has encountered an HCA heatmap, Figure 2 provides a primer and template for how to read the full HCA heatmap, with each major section described. Annotations are dichotomous variables, or, for categorical variables, the majority group is the reference group. We focus on education outcomes provided in ELS:2002 (dropout, SAT/ACT, 2-year degree, 4-year degree, STEM degree, STEM occupation by age 26, bachelor degree or above), as well as demographic and context variables (female, private or Catholic school with public as the reference group, suburban, urban, or rural; African American, Hispanic, or Asian with White as the reference group, if the student ever received financial aid in post-secondary school, standardized SAT score, SES, and overall GPA for all courses). Following the recommendations from the cluster analysis research literature (Weinstein, 2008), datapoints within the heatmap are color coded with gradients of blue (low z-scored “colder” grades) to red (high z-scored “hotter” grades), which also makes the heatmap accessible to people who are colorblind, such as one of the authors.

RESULTS:

As visual data analytics, the HCA heatmap patterns and visualizes the longitudinal grade trajectories of n=14,290 students across 6,728,920 individual data points. Figure 3 plots the full HCA heatmap. These results are a deep description of the lived experiences of over ten thousand students in American schools from grade 9 onward in each subject and grade level, linked to persistence, educational outcomes, and context and demographics. To help read the figure, as noted in the recent research on the application of data science visualization practices

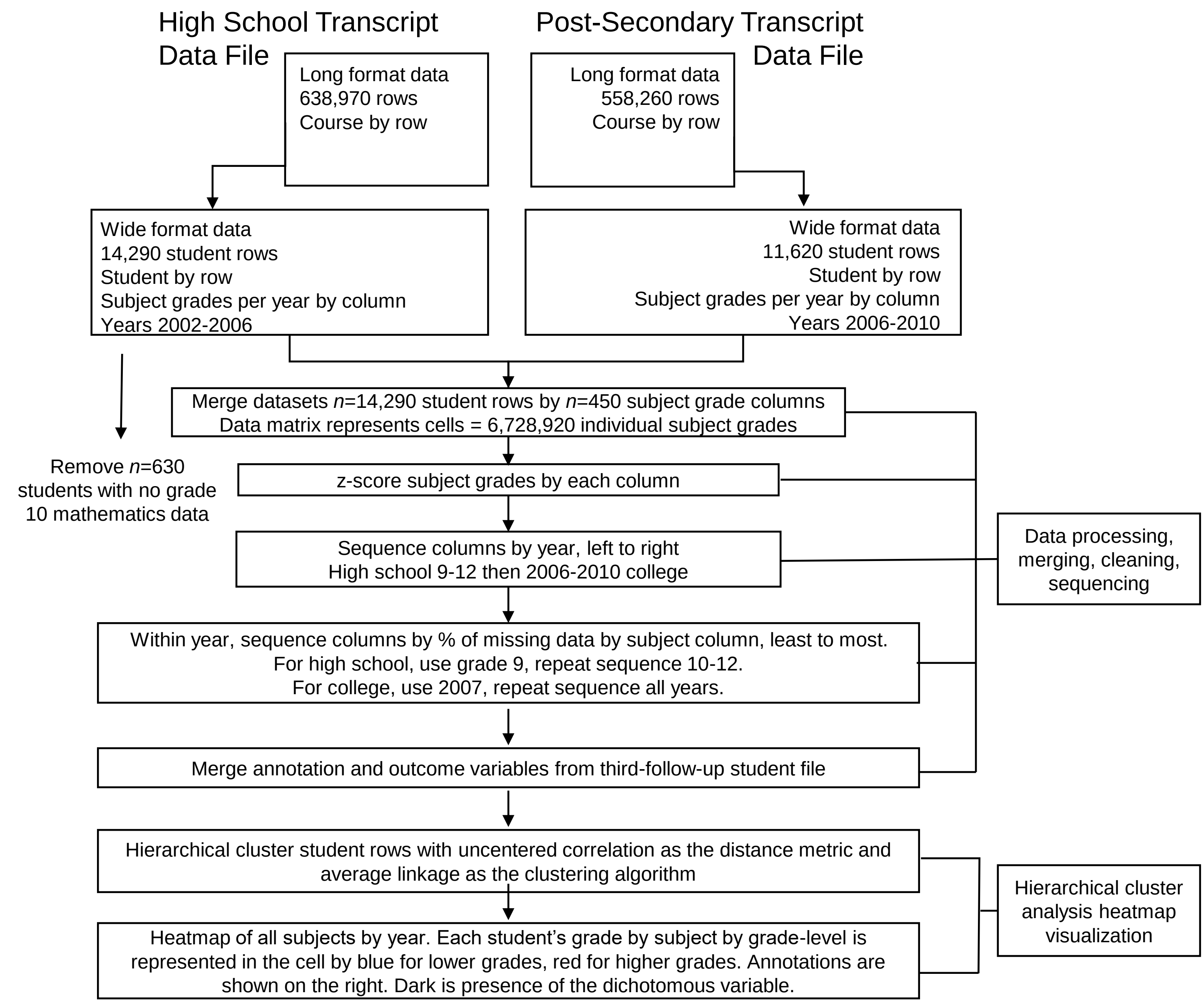


Figure 1: Overview of data processing and analysis design.

to education (Krumm & Bowers, 2022), we encourage the reader to "zoom in and out" of Figure 3, exploring specific columns, rows, or groups through zooming in to the high resolution figure, or zooming out. Additionally, for the annotations on the right, it can help the reader to hold a piece of blank paper up against the figure, revealing each annotation column one at time, left to right, to see the patterns. For example, covering all of the annotations except for the first left-most annotation for high school dropout reveals a quite stark pattern between the labeled Cluster 1 and Cluster 2. Additionally, it can be helpful to view the Appendix A order of the subjects for the columns for each grade-level. Of note, as discussed above, the HCA Heatmap in Figure 3 is a descriptive visual exploratory data analysis. Here, each row starts as its own cluster, with similar data patterns empirically being paired and agglomerated in a hierarchical fashion using the algorithm noted in the methods, such that the order of the rows in Figure 3, and by extension the overall clusters, are determined by the agglomeration algorithm given the distance metric (see methods).

Our results point to seven main findings. First, as evidenced by the cluster tree on the left of Figure 3, we find two main clusters color coded green (Cluster 1) and maroon (Cluster 2) of students who in general received high grades (red) (Figure 3, Cluster 1, bottom), and graduated high school and went to post-secondary school, in comparison to students who in general received average or lower grades (grey and blue) (Figure 3, Cluster 2, top) and who make up the majority of students who dropped out of high school (Figure 3, first annotation column, right, compare Cluster 1 to Cluster 2). This result replicates and extends Bowers

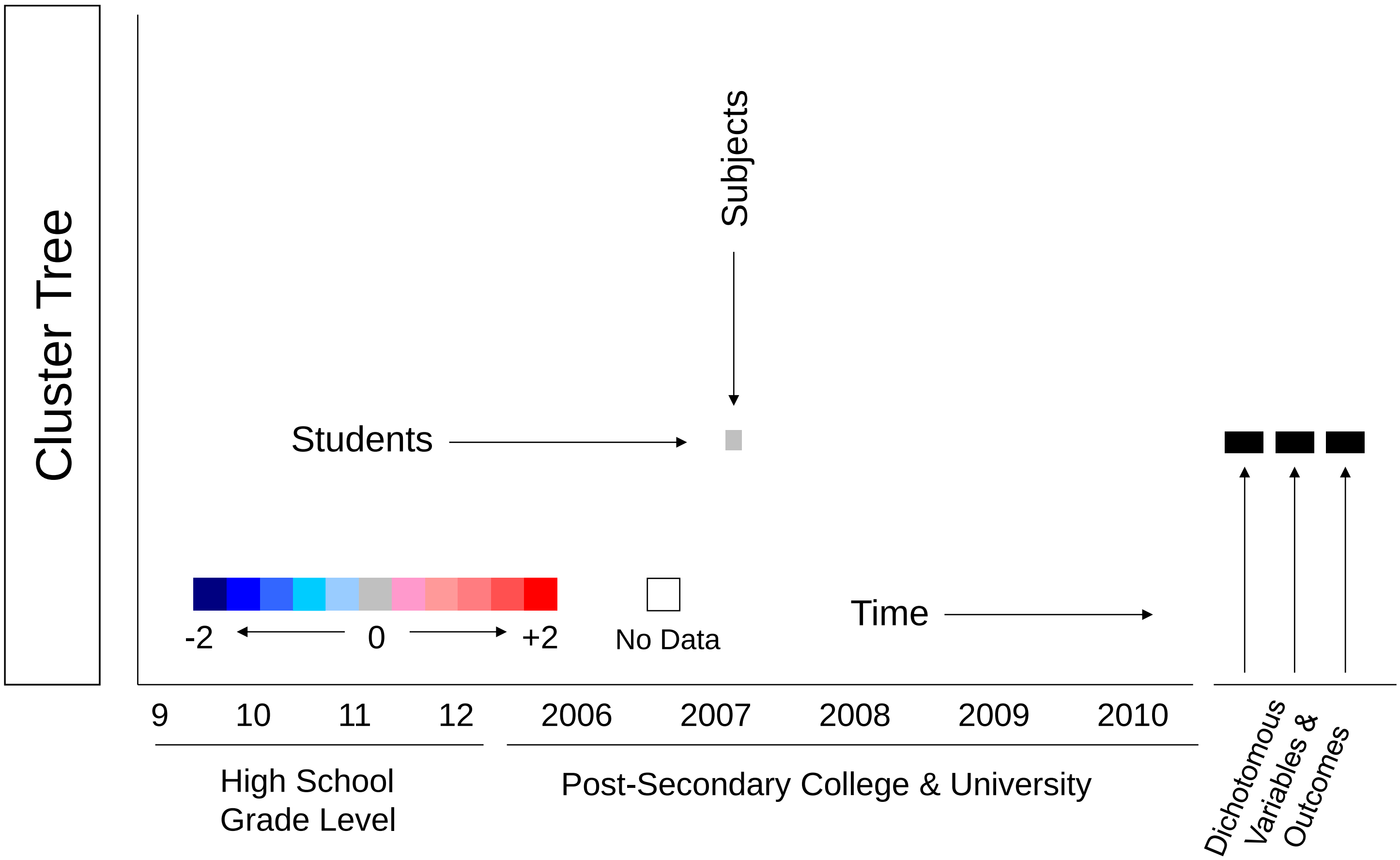


Figure 2: *Template for cluster analysis heatmap*. A cluster analysis heatmap contains multiple objects which represent the analysis results. The x-axis in our example is ordered in time, starting on the left with grades 9-12 of high school (secondary school), and continuing on through years in post-secondary college and university courses from 2006 through 2010, with 2006 as a year of possible dual-enrollment in both high school and college. The y-axis provides the cluster tree which graphically represents the similarity or dissimilarity between rows, with more similar rows clustered together and represented by shorter horizontal lines in the cluster tree. Thus, students are rows, and columns in this example are each course subject within each year. The list and order of course subjects in high school and post-secondary are listed in Appendix A. Any individual z-scored course grade for a student is represented by colors in a heatmap from higher grades (red) to lower grades (blue) with grey as the mean, and white representing no data. Annotations of dichotomous variables are provided on the right of the plot as dark bars for when the student has that variable.

(2010) findings to a large national sample that extends from high school through post-secondary school.

Second, within each grade level block of columns, the more contiguous columns to the left are the core subjects for each grade level, as each subject column in each grade level (year) is ordered left to right from more core-subjects to non-core subjects (see Appendix A), so the courses that students enroll in (and thus receive a grade in) are located to the left within each grade-level column block in the heatmap (Figure 3 center) from blue to grey to red (low to high grades), where white represents no data. Thus, Figure 3 visualizes the full course-taking and curriculum patterns for a large national sample, visualizing core-subject taking patterns in early high school, which then changes over time into more diverse subject course taking by the end of high school and into post-secondary school.

Third, in general, past performance predicts future performance, with some exceptions. The grades that students receive in grade 9 are generally similar across their time in high school and college, and grades received in one subject, in general, are similar to the grades received across all subjects, with some exceptions. For example, students in Cluster 1A (Figure 3, bottom), have a pattern of high grades throughout high school and college, graduate from high school, and go on to obtain 4-year post-secondary degrees, and have high rates of STEM degrees and STEM occupations at age 26. Cluster 1B students receive generally high grades throughout high school, except for the single blue column in grades 9, 10, and 11, for a select set of students, which is a low grade for citizenship (Figure 3 left, zoom in on the rows next to the B). These low citizenship graded students more often were in private and Catholic schools (Figure 3, annotations right), but otherwise resemble the patterns and

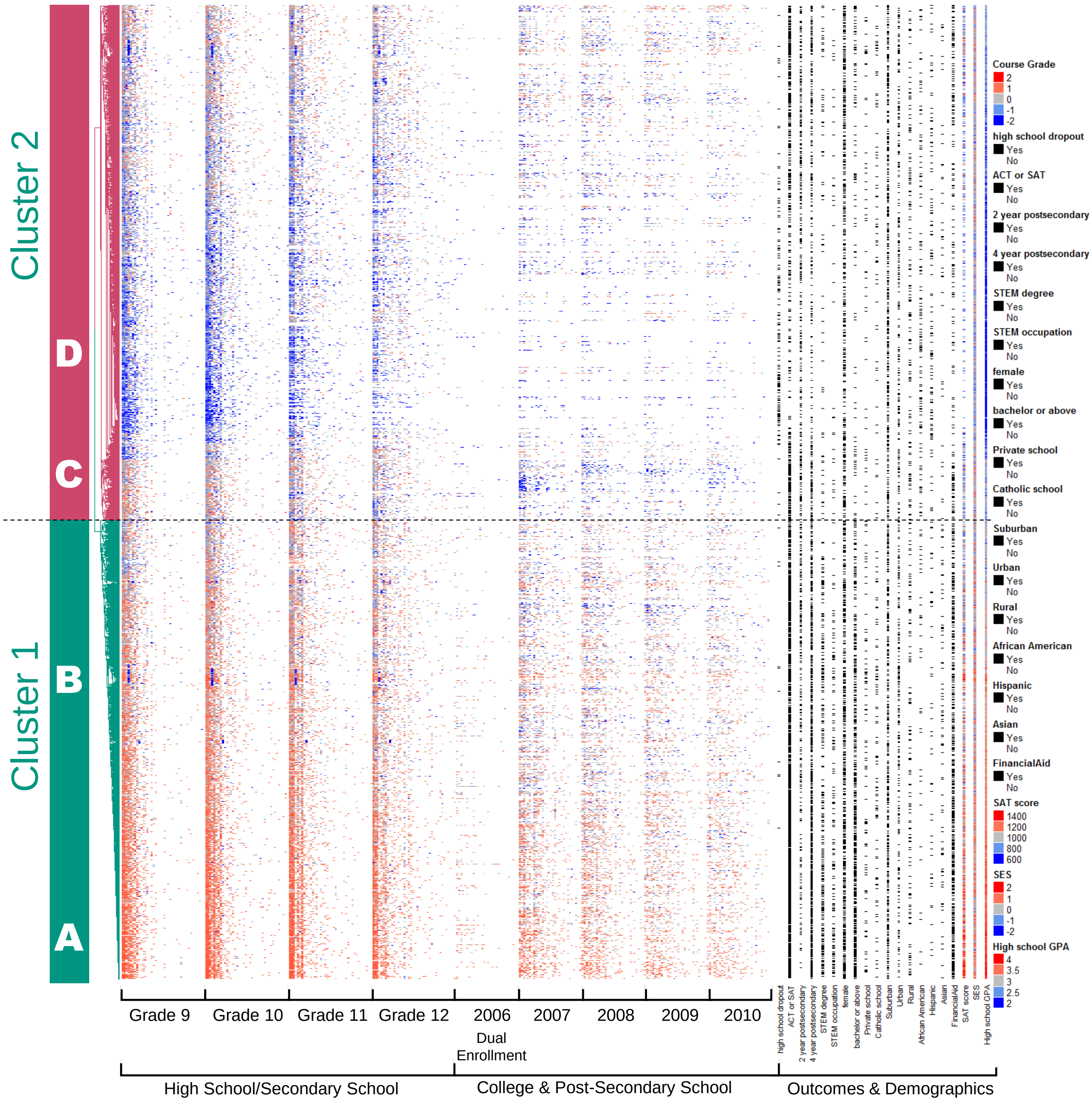


Figure 3: Hierarchical Cluster Analysis Heatmap of n=14,290 students (rows) by longitudinal course subject grades in each course in each grade level from grade 9 through four years of college. Two main clusters are identified, cluster 1 (left: green bar and cluster tree) and cluster 2 (left: maroon bar and cluster tree). Students in Cluster 1 have generally high grades (center: red) in high school and college. Cluster 1 students rarely drop out (right: annotations), do take the ACT or SAT, graduate with a two or four year degree, graduate with a STEM degree, and are more often in a STEM career by age 26. The highest performing students are in subcluster A. Subcluster B is one of the few patterns of students who do well in all classes in their grades (center: red) except one course each year of high school (blue) citizenship. Students in Cluster 2 (top) account for almost all of the high school dropouts, and do not generally have strong outcomes. Demographic annotations (far right) indicate that students in Cluster 2 are generally from historically underserved communities and contexts. Subcluster C includes students who in high school have slightly above average grades (light red) then lower grades in college (blue). Subcluster D students are the lowest performing students in the analysis (blue), dropping out most often (first annotation column).

outcomes of Cluster 1. A central finding from this analysis thus is that other than the citizenship columns (column 4 in each grade), the supposition in the research literature that there is a large set of students who do well in non-core subjects, such as art and music, but that there are consistent core-subject gateway courses such as algebra and mathematics that put the students off track (column 2 in each grade) (Allensworth, Nomi, Montgomery, & Lee, 2009; Carolan & Matthews, 2015; Gamoran & Hannigan, 2000) is not apparent in Figure 3. If this were the case, then other vertical bands of blue within individual columns would be obvious. Thus, when considering the full socio-curricular ecosystem visualized and mapped through an HCA heatmap, this analysis does not support the "gateway class" literature that focuses exclusively on courses such as algebra I. No study previously has described all individual subject course grades across this length of time for this many students in a national sample. Given that this is the first time such data at the national level has been visualized, it is a striking result that across the heatmap, especially in early high school, other than the citizenship course, student performance in one course generally is similar to their performance in all other courses. Thus, this finding supports the more holistic and individual-centered socio-curricular ecosystem framework described above (Frank et al., 2008; Heck et al., 2004), rather than a focus on overall averages or individual course subject performance, as students do not appear to operate within individual subject courses in isolation.

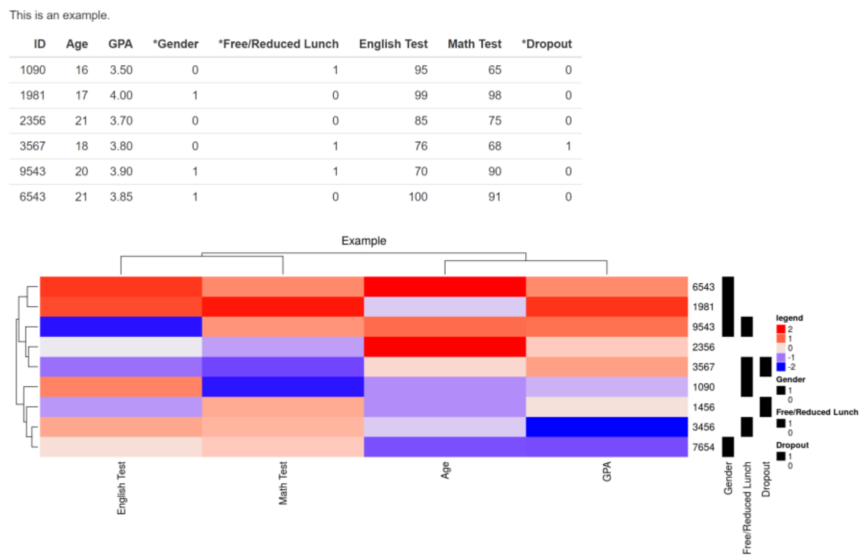
Fourth, the longstanding issues of segregation and advantage in the USA provided to students from high SES families in comparison to low SES families are visualized here in Figure 3. The second-to-last annotation column on the far right is color coded as a heatmap itself from high SES (red) to low SES (blue). Students in Cluster 1 who receive high grades and graduate from high school and college have generally higher SES, while students in Cluster 2 who receive lower grades in general and students with the lowest grades (Figure 3 Cluster 2D, blue) are from families with lower SES as well as more often historically disadvantaged racial and ethnic groups. However, there are exceptions with low SES students in cluster 1 (Figure 3 bottom right, note instances of blue in the SES annotation) and high SES students in cluster 2 (Figure 3 top right, note instances of red in the SES annotation). Thus, as a descriptive visual data analytic technique, rather than focus on averages which would mask this interesting heterogeneity, an HCA heatmap of this type provides the entire set of data for visual inspection and interpretation by the reader, who can themselves trace the lived experiences of individual students, *as individuals*, throughout the course of the entire dataset, noting overall trends, but also exceptions.

Fifth, Cluster 2C is intriguing because the cluster tree indicates that these students are somewhat dissimilar from the other student patterns in Cluster 2 (Figure 3, left, longer horizontal cluster tree lines). Cluster 2C students in general have just above average grades across subjects in grades 9 and 10, but then begin to struggle with lower grades starting in grades 11 and 12. Strikingly, for a large number of students in Cluster 2C, they appear to struggle immediately when they start college in 2007 and then throughout the rest of their time there (Figure 3, note change to blue for 2007 reading left to right, Cluster 2C). This visualizes the struggle these students experience as they move from high school to college.

Sixth, in examining the height of the cluster trees (Figure 3 left), also referred to as a dendrogram in cluster analysis (Romesburg, 1984), the horizontal length of the cluster tree lines on the rows can be interpreted as the relative similarity between the rows, with longer lines indicating more dissimilarity in longitudinal student course taking and grades patterns. Overall similarity of longitudinal student course taking and grades is higher in Cluster 1, students with generally high grades longitudinally (Figure 3, bottom left, shorter horizontal lines in the cluster tree), while dissimilarity is higher across Cluster 2, students with generally lower grades longitudinally (Figure 3, top left, longer horizontal lines in the cluster tree). This finding should be interpreted with caution however, as Cluster 1 students are more likely more similar in their longitudinal pattern due to the ceiling effect of the top end of the grading scale.

And finally, seventh, in considering the EWI/EWS literature noted above, and the goal of applying visual data analytics to this domain to help to identify intervention points to support student persistence and success when they meet significant challenges in school, for the students in Cluster 2, especially for the students in Cluster 2D who drop out most often, their data pattern in the HCA heatmap in Figure 3 is clear and striking in grade 9. Our argument here is not that the HCA heatmap is predictive, but that it is *descriptive*. Knowing that a student's grades are below average is not enough, as evidenced by these students continuing to struggle throughout their time in schools. For many students in Cluster 2D, their grades get worse by grade 10 (deeper blue in comparison to grade 9), and by grade 11 white streaks show up as they drop out and their data becomes missing (represented by white as well as noted in the dropout annotation column, Figure 3 center and right). While this result replicates previous research that has demonstrated that one of the most accurate predictors of dropping out is non-cumulative course grades from grade 9 and 10, especially a pattern of declining grades (Bowers & Sprott, 2012a; Bowers et al., 2013), here we show that this low grading pattern exists longitudinally across all subjects, both core subjects and non-core subjects. This result supports previous findings that there is a typology of students who drop out who need quite different types of interventions, from a reconnection to schooling as an overall good (such as here for Cluster 2D), to academic tutoring, to course enrollment and transcript audits to ensure that students are not unexpectedly missing required course credits (Bowers & Sprott, 2012b; Freeman & Simonsen, 2015; Sansone, 2019). Surfacing these patterns as early as possible across all of the students in a district, region, state, or nationally, provides a novel means for educators to identify individual students, subjects, courses, and grade levels in which students may need significant supports. We argue here that HCA heatmaps of this type, clustering enrollments and grades through time, are a useful means to visualize and pinpoint the challenge of individuals as well as groups and clusters of students for potential interventions and supports.

Panel A:



Panel B:

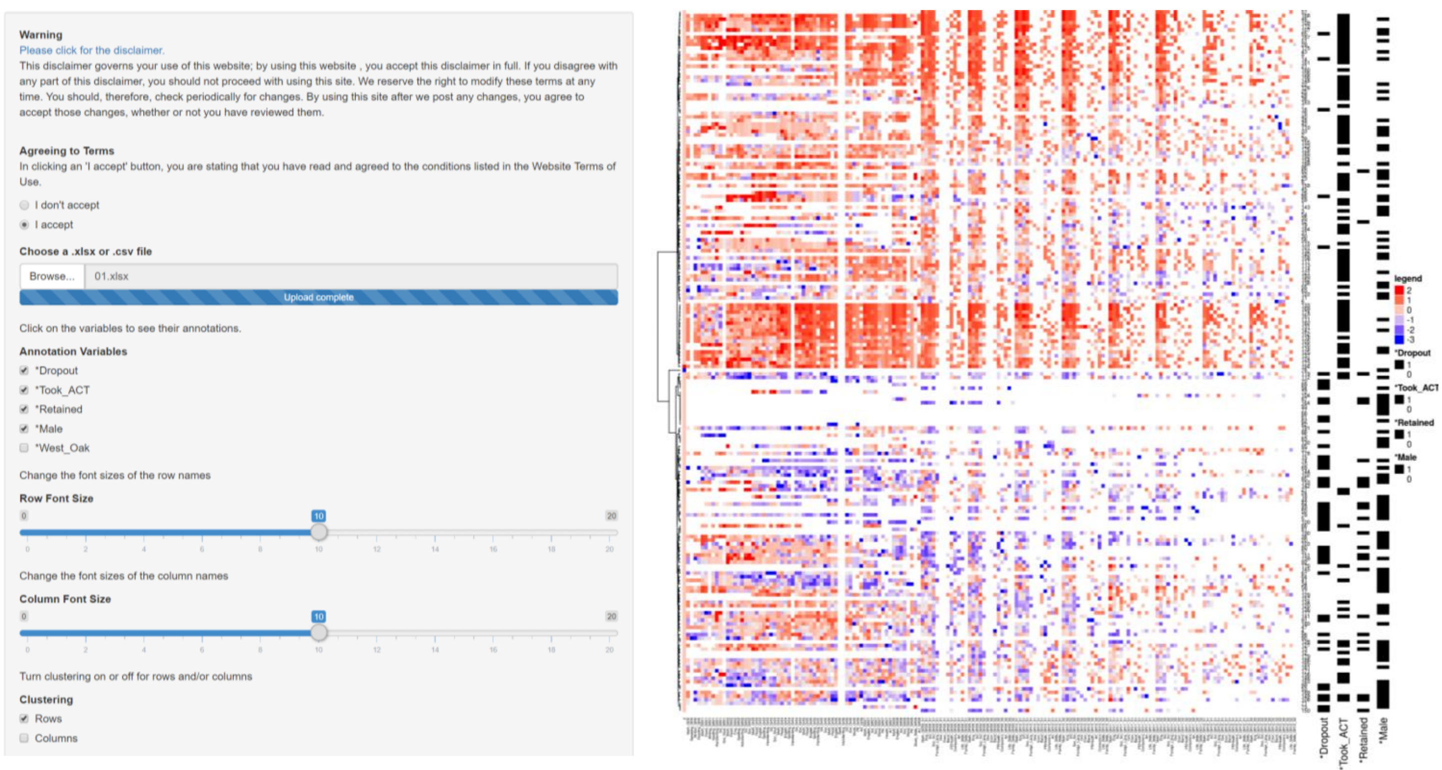


Figure 4: *Visual Data Analytics R Shiny Website Screenshot of Cluster Analysis Heatmap for Education*. To provide practitioners the ability to create HCA cluster analysis heatmaps without the need to access the R code, we created an open access website using R Shiny. The intention of the website is to allow practitioners to generate HCA heatmaps easily to the recommended specifications and annotations. To upload the data and manipulate the visualization, users must first agree to terms and privacy warning, then can upload their data as either a .csv (comma delimited) or .xlxs (Microsoft Excel) files. Data must be numeric, with missing data either as blanks, ".", or "NA". If annotation data is included, the user must add an asterisk as the first character in the column label for that variable and the variable will be automatically detected as an annotation variable. Once data is uploaded, it will be displayed (Panel A: top) for inspection. A heatmap of the data is automatically generated following previous recommendations from the literature (Bowers, 2010) with each column z-scored, with higher intensity of red indicating higher levels of the variable, blue lower levels, grey as the average, and white as no data (Panel A: bottom). Annotation variables are indicated on the left and can be included with checkboxes. Hierarchical cluster analysis with uncentered correlation as the distance metric and average linkage as the clustering algorithm for either the rows, columns or both are provided as checkboxes, (Panel B, left). Row and column label font sizes can be controlled with sliders. In Panel B we have replicated the results from Bowers (2010) by clustering the dataset used in the 2010 study, representing n=188 student

longitudinal grade patterns in all subjects from grades K-12. Annotations include if the student dropped out, took the ACT college entrance exam, was retained in any grade level, or is male. HCA clustering is selected for the rows only. The HCA heatmap in Panel B replicates the findings from Bowers (2010) with two main clusters, one of generally high grades (red), graduation, and taking the ACT (Panel B: right top cluster), and one of generally low grades (blue), dropping out, retention, and not taking the ACT (Panel B: right lower cluster). While we provide the example of student longitudinal grade data, any data can be uploaded and clustered on rows or columns, or both, providing a new visual data analytic tool to the research and practice community.

To make the HCA heatmap technique accessible to EWI/EWS education practitioners, we provide the full open source code in R (https://doi.org/10.7916/cqvn-9t71) within an online web browser environment in the open source application R Shiny (https://ohrice.shinyapps.io/Heatmap) which is also available for download to run locally and encrypted. Figure 4 details the R Shiny application that provides the visual data analytic tool of HCA heatmaps to anyone through uploading a data file and visualizing the heatmap. Here we demonstrate the tool with mock data (Figure 4, Panel A), as well as the longitudinal grade data from Bowers (2010) of *n*=188 student grades across all subjects K-12 with outcome annotations (Figure 4, Panel B), replicating the HCA heatmap results from that study using data from two small school districts. Note that education practitioners can include student IDs in the data as an option, which then will be displayed on the right of the HCA heatmap between the heatmap and the annotations. Our goal is to support the use of HCA heatmaps in education, through providing the code and an in-browser application, as the cluster analysis heatmap is well-suited to pattern longitudinal grades data as shown here, but also can pattern and visualize many different types of data, both on the rows and the columns.

DISCUSSION:

In this study we demonstrate the usefulness of HCA heatmaps as a descriptive visual data analytic technique for EWI/EWS to describe, pattern, and visualize the individual experiences of more than ten of thousand students throughout each subject and grade level from grade 9 through college. For evidence-based decision making in schools (Mandinach & Schildkamp, 2021) and the growing literature on EWI/EWS (Balfanz & Byrnes, 2019; Bowers, 2021b), HCA heatmaps are a useful means to surface important patterns of student success and challenge throughout their time in schools. Here, we replicate and extend the previous literature on the usefulness of patterns of teacher-assigned grades across subjects to inform the work of EWI/EWS (Bowers & Sprott, 2012a, 2012b; Bowers et al., 2013; Brookhart et al., 2016), and visualize the complete transcript records of all of the students in the dataset as a socio-curricular ecosystem (Frank et al., 2008; Heck et al., 2004).

This study is novel and significant in three main ways. First, we find strong evidence for the socio-curricular ecosystem perspective of student course taking and performance, and little evidence that there are specific gateway courses or subjects that limit or enable student long-term outcomes. Rather, the relationships we visualize here across more than 14,000 students and more than 6 million individual course subject grades representing and visualizing the entire enrollment and grading history for each student, indicate that students appear to be supported or challenged in similar ways across the entire curriculum, as a socio-curricular ecosystem. Given these findings, we argue that in working to promote student persistence and success in the system, focusing on individual course subject achievement may be naïve. Research traditionally ignores the full slate of subjects and courses, focusing instead on the researchers' or policymakers' interests, such as algebra, science, or reading and writing. Rarely are all courses considered as an ecosystem (Bowers, 2011; Frank et al., 2008; Heck et al., 2004), including social studies, the arts, career and technical education, and the entire range of subjects offered in the US curriculum (Powell, Farrar, & Cohen, 1985). Our findings indicate that students generally perform well or are challenged in similar ways across subjects and courses. In contrast to the socio-curricular ecosystem framework, one might alternatively conclude from these findings, however, that focusing on a specific course or small set of courses is ultimately a good approximation of overall performance. We argue that this alternative interpretation of the results would be confusing an individual symptom with the much larger system. Focusing on a specific course, such as algebra 1 for example, would not take into account our central finding from the socio-curricular ecosystem framework that student performance as visualized and patterned here across all of their enrolled courses and grades does not appear to be course specific, and so focusing on a specific course seems to perhaps miss the forest for the trees. Thus, we encourage future research to further consider how individual courses each contribute to the overall socio-curricular ecosystem, and how perhaps interventions that help support overall student success working to address systematic issues across all courses as a system itself, as postulated in the previous research on the socio-curricular ecosystem (Frank et al., 2008; Heck et al., 2004), could help support findings from individual subject-specific course interventions.

Second, we replicate and extend Bowers (2010), here at the national level, finding two large clusters that are strongly related to high school graduation and long term post-secondary outcomes. Third, student success or challenge with schooling is obvious in grade 9. For students in cluster 2, visualizing only their individual grades in each course and subject, data that are already collected in schools and reported to students on report cards, indicates challenge with schooling across subjects at the start of high school. The HCA heatmap method visualizes this issue, surfacing and highlighting these student's lived experiences, and providing an opportunity to make these previously unseen patterns knowable and therefore actionable.

Visualizing the Socio-Curricular "Map" of Each Individual Student Life Course

We envision this technique being useful beyond the ELS:2002 transcript dataset for individual schools, districts, and state and regional education systems, as education data scientists (Agasisti & Bowers, 2017; Bowers, 2017, 2021a, in press; Bowers et al., 2019; Bowers & Krumm, 2021; Bowers et al., 2016; Krumm & Bowers, 2022; Piety et al., 2014; Piety & Pea, 2018) work to describe the diverse array of data that exists within and across schools. While we have detailed the major patterns displayed in the HCA heatmap, as a description of every student in the dataset that includes the patterns of grades across each subject for many years of schooling, there are many more patterns to explore that are beyond the scope of the present study. As a descriptive visual data analytic technique (Bienkowski et al., 2012), the reader's ability to form and then test hypotheses immediately by looking at the figure is a strength of cluster analysis heatmaps across multiple domains (Bowers, 2010; Wilkinson & Friendly, 2009), and demonstrated here in education. Considered from the human-computer interaction design theory framework (Preece et al., 1994) in which map-like information helps the reader see the entire dataspace at once, examine both usual and alternative routes and edge cases, and previously uncharted or rarely examined areas of interest alongside the well-known (Verplank, 2003), we argue here that cluster analysis heatmaps are a useful addition to the education research data display lexicon. This type of visualization helps to complement both in-depth descriptions from qualitative research as well as the traditional quantitative education research focus on summaries, averages, and regression statistics, and provide deep, thick, and rich descriptions of every individual in the dataset, mirroring many of the goals of qualitative research (Denzin & Lincoln, 2000), visualizing the lived experiences of every student through their time in the schooling system.

Issues of Equity, Ethics, and Privacy in Visual Data Analytics

We also provide the open access code and digital tool as a means to empower data users and education data scientists in their organizations and communities (Agasisti & Bowers, 2017; Bowers, 2021a, 2021b; Bowers et al., 2019). Data science, machine learning, and pattern analytics has a problematic history across many social, governmental, and business domains, as the historic data used to train machine learning algorithms will ensure that those algorithms reproduce longstanding problematic issues of segregation, racism, and historical disadvantage of the poor, especially in health and education (Benjamin, 2019; Bowers, 2021b, in press; Bowers et al., 2022; Hawn Nelson et al., 2020). As noted by Benjamin (2019) in relation to health care:

> Data used to train automated systems are typically historic and, in the context of health care, this history entails segregated hospital facilities, racist medical curricula, and unequal insurance structures, among other factors. Yet many industries and organizations well beyond health care are incorporating automated tools, from education and banking to policing and housing, with the promise that algorithmic decisions are less biased than their human counterpart (p.422).

In education, and USA schooling systems in particular, curricular options through tracking and segregation have inequitably limited access to these resources to students from historically underserved communities (Oakes, 2005; Reardon, 2019). A clear danger exists when applying data science and machine learning to education, in that the machines will learn and replicate past problematic systems (Bowers, 2021b, in press; Fischer et al., 2020). Here, our goal is to provide an alternative to hidden algorithms, summaries, averages, and the exclusive (and biased) narrative of an author of a study or visualization that relies exclusively on averages and regression coefficients, and instead surface and describe patterns visualizing all of the data, and provide open and accessible visual data analytic tools that allow the reader to see the author's summary while coming to their own conclusions as all of the patterns and data are visualized together.

This intention throughout the present study on moving away from an exclusive reliance on summary statistics such as averages and regression coefficients, and instead provide deeper descriptions of the patterns of individuals and their individual data points that represent each student's own journeys through the system, aligns with recent calls for a critical and equity-centered focus in data use throughout educational decision making (Mandinach & Schildkamp, 2021). For example, in discussing the intersection of equity and improvement science in service to historically underserved communities, Hinnant-Crawford highlights this focus on individuals throughout the process by noting "if we truly want to understand the nature of improvement driven by human behavior, we have to examine both the logic and the psychology that drive choices and decisions of individuals with agency within our organizations." (Hinnant-Crawford, 2020, p. 36). Similarly, in their report focused on data use in schools titled "Centering Racial Equity Throughout Data Integration", Hawn-Nelson et al. (2020) note that in comparison to problematic data practices of "attempting to describe individual experiences with aggregate or "whole population" data" (p.31) a positive aspect of data use that centers equity is the "whole-person view" in which:

> When data are integrated across multiple sources, we get a more holistic view of the experiences and outcomes of children, households, and families, supporting asset- (rather than deficit-) based approaches. Such views allow analysts to identify bright spots across communities, families, and individuals, and, ultimately, encourage investment in policies and programs that work. (Hawn Nelson et al., 2020, p. 12)

Indeed, visual data analytics of the type discussed throughout this study aligns with this research on equity-centered practices across data use and the curriculum that encourages a more holistic view that includes the individual within the larger system, and promotes a focus that goes beyond outcomes such as test scores and graduation rates, and includes opportunities and resources (Carter & Welner, 2013; Gutiérrez, 2012). For example, recent reports by the National Academy of Science, Engineering, and Mathematics (NASEM) on building educational equity indicator systems, have provided a framework and recommendations for school systems to move beyond an outcomes focus and include and report multiple indicators of equity (National Academies of Sciences Engineering and Medicine, 2019, 2020), of which curricular breadth, access to and enrollment in rigorous coursework, and performance in coursework, are three NASEM equity indicators that are informed through a socio-curricular framework that uses cluster analysis heatmaps to visualize, pattern, and describe the entire history of individual student course enrollments and grades.

Yet, while an individual perspective to this type of analytics is recommended across these reports, recent studies also note issues of ethics, privacy, and access when it comes to the use of big data analytics in education research and practice (Fischer et al., 2020; Hawn Nelson et al., 2020; National Academies of Sciences Engineering and Medicine, 2019; National Academy of Education, 2017), especially for education early warning systems (Bowers, 2021b). These issues are, of course, of concern for cluster analysis heatmaps of large sets of student data as the technique can pattern and visualize the individual data of thousands of students across millions of datapoints, such as here. Yet, this issue is not unique to the application of HCA heatmaps to education, as noted above, HCA heatmaps have a long history of being applied extensively throughout industries such as cancer biology and oncology, an industry with not dissimilar privacy and ethical concerns to education data, visualizing thousands of human cancer patient's data across all of their gene expression levels for thousands and tens of thousands of genes (Wilkinson & Friendly, 2009).

A central concern in analytics in education is to not only not release personally identifiable information that can be linked back to identified individuals, but that the data analysis and display of the results itself does not allow for the re-identification of de-identified individuals using the analysis reported (Fischer et al., 2020; National Academy of Education, 2017). Drawing on the lessons learned from the application of HCA heatmaps across the variety of fields with strong individual data privacy regulations noted throughout this study, HCA heatmaps provide a useful means to focus on individual-level data while maintaining data privacy in four main ways. First, at the most basic level, as demonstrated here and in the previous HCA heatmap studies in education noted above, each student's row in the heatmap is deidentified, such that there is no indication of a student number or unique identifier that would link back to individuals directly. However, for internal use within a school district or within an encrypted computing environment, student IDs can be listed and requested in the visualization, as detailed in the Shiny app and the demonstration R code appendices.

Second, each column of data in the heatmap is z-scored, thus obfuscating the original data and norming the resulting output within each column, which as noted in the methods, is a standard step in cluster analysis to ensure that the distances in the distance matrix are comparable on a standardized metric, with the added privacy benefit noted here of obfuscating the actual data. Additionally, this step is similar to grand mean centering in regression statistics, making the averages comparable across a wide range of courses and grade levels. While beyond the scope of the present study, future research may wish to investigate the difference in patterns without z-scoring.

Third, the actual data within the heatmap is never displayed as a number, but rather the heatmap visualization itself is a key component of maintaining privacy of the individuals, as by translating the individual datapoints into a heatmap that interpolates along a gradient from blue to red, each individual datapoint is not displayed as a number, rather a visual representation is displayed through the color of the heatmap. Thus, just as in fields with strong ethics and privacy regulations and concerns such as cancer biology, here in education the HCA heatmap provides a means to analyze and display individual data in a de-identified and obfuscated manner. And finally, fourth, as noted above, for district data analysts, we provide the full code in the appendices for the R shiny application and the tutorial in R, each of which can be run in an encrypted private computing environment locally within a school district on the analyst's computer.

Exploratory Visual Data Analytic Cluster Analysis Heatmaps as a Useful Tool in Education

However, taking the alternative view, one could ask in relation to our findings throughout this study on the application of HCA heatmaps to grades data, so what? There is already strong evidence across the literature that demonstrates that low grades are highly correlated on average with dropping out and high grades with persistence in school (Bowers et al., 2013; Brookhart et al., 2016). What is it here that HCA heatmaps add other than more work for the already beleaguered education data analyst (Agasisti & Bowers, 2017; Bowers, 2017; Bowers et al., 2019; Selwyn et al., 2021)? Indeed, current research using longitudinal inferential statistics, such as Growth Mixture Modeling (GMM), and non-cumulative GPA in the first three semesters of high school (grade 9 semester 1 and 2, and grade 10 semester 1), has been shown to be one of the most accurate predictors of dropping out (Bowers, 2021b; Bowers & Sprott, 2012a; Bowers et al., 2013; Coleman, 2021; Knowles, 2015). Additionally, mixture models, such as GMM and Latent Class Analysis (LCA), include significance tests on the longitudinal trajectories and multiple covariates, such that a probability distribution model is tested with the assumption that the data are a sample from a mixture from that distribution (Vermunt & Magidson, 2002). However, as noted by Vermunt and Magidson (2002), while mixture models of this type are similar to nonhierarchical cluster analysis, such as k-means, outcomes may be quite similar. As with k-means clustering, these models are a top-down *a priori* method to predict the groups. Additionally, these types of inferential models are restricted in the number and types of data and variables that can be included, especially when considering distributional assumptions, statistical power, interactions, and the like. In the present study, our focus is on socio-curricular theory and using a bottom-up map-like approach to describe and visualize the longitudinal journeys of each student and the relationships across these journeys in the dataset throughout their time in the system to augment and complement the path-like information provided through the inferential techniques described throughout the literature in this domain.

We posit that similar to the bioinformatic literature noted throughout this study, for the education research literature the answer to the "so what?" question is that even when research is quite specific for one variable that indicates a strong correlation *on average*, this average provides little in the way of being able to help educators take informed action to intervene in positive ways to help address specific student needs. In an average, the individual is lost. Visual data analytics (Bienkowski et al., 2012) in the tradition of Exploratory Data Analysis (EDA) (Behrens, 1997; Tukey, 1962, 1977) that address the socio-curricular ecosystem that students experience (Frank et al., 2008; Heck et al., 2004), of the type demonstrated here of HCA heatmaps, use the same data and add the ability to see the individual within the larger pattern. The addition of HCA heatmaps as a tool for education data analysts (Bowers, 2021b; Bowers et al., 2019; Bowers & Krumm, 2021), both in districts and in research and policy, will simultaneously help broaden the view across the disaggregated data patterns by displaying all of the data at once, and allow for the life course and lived experience of individuals (Alexander et al., 2001; Pallas, 1989, 1993) to remain present in the analysis. We encourage future research and practice to include HCA heatmaps not as a replacement for other visualization, description, and data summaries, but to augment and deepen the warrants for their claims. Indeed, the recent research on evidence-based improvement and data driven decision making practice in schools and districts demonstrates that data that are specific and individual to teacher practice can have the most influence on instructional improvement (Gerzon, 2015; Halverson, 2010; Mandinach & Schildkamp, 2021).

While outside the scope of the present study the HCA heatmap is, as with bioinformatics, just one part of a suite of evidence to warrant claims that data patterns across the heatmaps are predictive of overall outcomes. When we started this study, given the national level of the dataset, the authors had assumed that we would identify strong "blocks" of blue or red in high school or college that correlated with outcomes, such as for specific students struggling in core subjects but doing well otherwise, or vice versa. However, we were surprised that overall, other than the citizenship exception, we found little evidence for these types of "gateway course" subject specific patterns, and rather Cluster 1 students generally receive high grades and go to college, and Cluster 2 students generally receive lower grades and drop out much more often, replicating the much smaller district-embedded study of Bowers (2010). Nevertheless, we encourage future research to further explore the heterogeneity represented across the HCA heatmap, as there are intriguing yet small patterns that may warrant further study.

Towards More Quantitative Large-Scale Visual Descriptive Research in Education

As noted across the Exploratory Data Analysis (EDA), visual data analytics, and education data science research literature (Agasisti & Bowers, 2017; Bienkowski et al., 2012; Bowers, in press; Bowers et al., 2022; Bowers & Krumm, 2021; Krumm & Bowers, 2022; Krumm et al., 2018; Piety et al., 2014; Tukey, 1962, 1977), visualizing and describing the complex relationships across large-scale datasets is an important contribution for a research and practice field (Gelman & Vehtari, 2021) such as education and education early warning systems (Bowers, 2021b). Concurrently, across the psychology research literature, there is the current discussion about the "replication crisis" in which many past experiments tested with traditional individual variable hypothesis tests fail to replicate (Yarkoni, 2022). In education research, and especially recent large scale and randomized controlled EWI/EWS studies (Faria et al., 2017; Mac Iver et al., 2019; Sepanik et al., 2021), as well as multiple large-scale education data use studies (Farley-Ripple, Jennings, & Jennings, 2021; Gleason et al., 2019; Grabarek & Kallemeyn, 2020; Wayman, Shaw, & Cho, 2017), beyond select high performing case studies, researchers have failed to identify significant relationships between early warning system use and student persistence, achievement, and degree completion on average. Traditionally in education research, quantitative studies often rely on individual variable p-value significance hypothesis testing. We argue here throughout the present study, that perhaps the single variable average summary statistics focus throughout much of quantitative education research is not serving the research and practice domain fully, and perhaps could be augmented and complemented by descriptive visual data analytics, such as the HCA Heatmap of the socio-curricular ecosystem discussed throughout the present study. Recently, one solution proposed for the replication crisis has been to "take descriptive research more seriously" (p.17) (Yarkoni, 2022, p. 17). Specifically, as stated by Yarkoni (2022):

> Traditionally, purely descriptive research—where researchers seek to characterize and explore relationships between measured variables without imputing causal explanations… is looked down on in many areas of psychology. This stigma discourages modesty, inhibits careful characterization of phenomena, and often leads to premature and overconfident efforts to assess simplistic theories that are hopelessly disconnected from the complexity of the real world… Acknowledging the value of empirical studies that do nothing more than carefully describe the relationships between a bunch of variables under a wide range of conditions would go some ways towards atoning for our unreasonable obsession with oversimplified causal explanations. We know that a large-scale shift in expectations regarding the utility of careful descriptive work is possible, because other fields have undergone such a transition... Perhaps most notably, in statistical genetics, the small-sample candidate gene studies that made regular headlines in the 1990s… have all but disappeared in favor of massive genome-wide association studies (GWAS) involving hundreds of thousands of subjects (Nagel et al., 2018; Savage et al., 2018; Wray et al., 2018). The latter are now considered the gold standard even in cases where they do little more than descriptively identify novel statistical associations between gene variants and behavior. In much of statistical genetics, at least, researchers seem to have accepted that the world is causally complicated, and attempting to obtain a reasonable descriptive characterization of some small part of it is a

perfectly valid reason to conduct large, expensive empirical studies (p.17) (Yarkoni, 2022).

We concur with Yarkoni's argument, as this encouragement to take descriptive research more seriously is a core component of the usefulness of HCA heatmaps applied to large-scale quantitative education data. In the present study, using a visual data analytic technique applied to the socio-curricular theory of the student life-course, rather than find single individual courses and subjects that seem to gate or limit students, in viewing the entire map of the journey of more than ten thousand students across more than six million individual subject grades and enrollments which includes a wide variety of courses, subjects, and grade levels, we provide a description of the relationships across this highly varied and multidimensional data. This technique provides an opportunity to investigate these relationships both overall across the entire dataset and zoom in on both clusters and individual student journeys through the system. Given the lack of findings to date in the EWI/EWS research literature, we argue that this technique of describing large scale relationships across such a wide variety of data while retaining the individual data story, is a useful complement to the vast majority of the EWI/EWS research that focuses on average summary statistics of specific variables.

Conclusion:

Here we applied HCA heatmaps to a large national level dataset, demonstrated the utility of visual data analytics, and then replicated the results from two small districts to relate directly to practice. We encourage future researchers to further apply HCA heatmaps to other school districts, regions, states, and nations, to visualize and describe the longitudinal cluster patterns across the wide range of data collected in schools beyond course grades, such as attendance, standardized test scores, benchmark scores, satisfaction, interest, engagement, and more. We selected a large national dataset here to demonstrate the utility of the method with large datasets that represent data that is not unlike data that is collected on a regular basis across school districts world over. Additionally, as the dataset is available with a restricted data license, rather than a district-specific dataset, future research can access the data from NCES to replicate and extend the study. As with the use of HCA heatmaps across a wide range of big data domains noted above, describing the individual patterns and displaying every data point, rather than obfuscating and hiding the heterogeneity behind averages and standard deviations, allows the reader, practitioners, and researchers to see the individuals within the data and identify patterns and relationships in large multidimensional data matrices that would go undetected otherwise. We encourage future researchers and education data analysts to consider HCA heatmaps as a useful tool to address these goals.

Note: Online Supplement Appendices:

Online Supplement 1 Appendix for Bowers, Zhao, and Ho (2023) Towards Hierarchical Cluster Analysis Heatmaps: R Shiny application code: https://doi.org/10.7916/cqvn-9t71

Online Supplement 2 Appendix for Bowers, Zhao, and Ho (2023) Towards Hierarchical Cluster Analysis Heatmaps: R Tutorial and code: https://doi.org/10.7916/r1mg-yn37

REFERENCES:

Adams, E., Trojanowski, A., Davis, J., Agramonte, F., Hazle Bussey, L., Giarrizzo, A., & Krumm, A. E. (2021). Collaborative Data Visualization: A Process for Improving Data Use in Schools. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists.* (pp. 266-280). New York, NY: Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Agasisti, T., & Bowers, A. J. (2017). Data Analytics and Decision-Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions. In G. Johnes, J. Johnes, T. Agasisti, & L. López-Torres (Eds.), *Handbook on the Economics of Education* (pp. 184-210). Cheltenham, UK: Edward Elgar Publishing. https://doi.org/10.7916/D8PR95T2

Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, *103*(5), 760-822.

Allensworth, E. M. (2013). The Use of Ninth-Grade Early Warning Indicators to Improve Chicago Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 68-83. doi:10.1080/10824669.2013.745181

Allensworth, E. M., Nagaoka, J., & Johnson, D. W. (2018). *High School Graduation and College Readiness Indicator Systems: What We Know, What We Need to Know*. Retrieved from Chicago, IL: https://consortium.uchicago.edu/sites/default/files/publications/High%20School%20Graduation%20and%20College-April2018-Consortium.pdf

Allensworth, E. M., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College Preparatory Curriculum for All: Academic Consequences of Requiring Algebra and English I for Ninth Graders in Chicago. *Educational Evaluation and Policy Analysis*, *31*(4), 367-391. doi:10.3102/0162373709343471

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

Baker, R. S., Berning, A. W., Gowda, S. M., Zhang, S., & Hawn, A. (2020). Predicting K-12 Dropout. *Journal of Education for Students Placed at Risk (JESPAR)*, *25*(1), 28-54. doi:10.1080/10824669.2019.1670065

Balfanz, R., & Byrnes, V. (2019). Early Warning Indicators and Intervention Systems: State of the Field. In J. A. Fredricks, A. L. Reschly, & S. L. Christenson (Eds.), *Handbook of Student Engagement Interventions* (pp. 45-55): Academic Press.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131-160. doi:10.1037/1082-989X.2.2.131

Behrens, J. T., DiCerbo, K. E., Yel, N., & Levy, R. (2012). Exploratory Data Analysis. In *Handbook of Psychology, Second Edition*.

Benjamin, R. (2019). Assessing risk, automating racism. *Science, 366*(6464), 421-422. doi:10.1126/science.aaz3873

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. Retrieved from Washington, DC: http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

Bland, M. J., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310. doi:https://doi.org/10.1016/S0140-6736(86)90837-8

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration, 47*(5), 609-629. doi:10.1108/09578230910981080

Bowers, A. J. (2010). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment Research and Evaluation, 15*(7), 1-18. https://doi.org/10.7275/r4zq-9c31

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159. doi:10.1080/13803611.2011.597112

Bowers, A. J. (2017). Quantitative Research Methods Training in Education Leadership and Administration Preparation Programs as Disciplined Inquiry for Building School Improvement Capacity. *Journal of Research on Leadership Education, 12*(1), 72 - 96. doi:10.1177/1942775116659462

Bowers, A. J. (2019). Towards Measures of Different and Useful Aspects of Schooling: Why Schools Need Both Teacher Assigned Grades and Standardized Assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom Assessment as Educational Measurement, National Council on Measurement in Education (NCME) Book Series* (pp. 209-223). New York: Routledge.

Bowers, A. J. (2021a). Dashboards, Data Use, and Decision-making – A Data Collaborative Workshop Bringing Together Educators and Data Scientists. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 1-36). New York, NY: Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Bowers, A. J. (2021b). Early Warning Systems and Indicators of Dropping Out of Upper Secondary School: The Emerging Role of Digital Technologies. In *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. Paris, France: Organisation for Economic Co-Operation and Development (OECD) Publishing. https://doi.org/10.1787/589b283f-en

Bowers, A. J. (in press). Data Science in Education Administration, Policy, and Practice. In L. Cohen-Vogel, J. Scott, & P. Youngs (Eds.), *AERA Handbook of Education Policy Research* (Vol. 2nd). Washington, DC: American Educational Research Association (AERA).

Bowers, A. J., Bang, A., Pan, Y., & Graves, K. E. (2019). *Education Leadership Data Analytics (ELDA): A White Paper Report on the 2018 ELDA Summit*. https://doi.org/10.7916/d8-31a0-pt97

Bowers, A. J., Choi, Y., Shi, H., Sun, F., Sun, H., Williams, J., & Weeks, S. (2022). *Mapping 16 Equity Indicators to the New York City Schools Public Datasets: Access, Opportunity, and Outcomes*. Retrieved from New York, NY: https://doi.org/10.7916/a143-aq05

Bowers, A. J., & Krumm, A. E. (2021). Supporting the Initial Work of Evidence-Based Improvement Cycles Through a Data-Intensive Partnership. *Information and Learning Sciences, 112*(9/10), 629-650. doi:10.1108/ILS-09-2020-0212

Bowers, A. J., Krumm, A. E., Feng, M., & Podkul, T. (2016). *Building a Data Analytics Partnership to Inform School Leadership Evidence-Based Improvement Cycles*. Paper presented at the Annual meeting of the American Educational Research Association, Washington, DC.

Bowers, A. J., Shoho, A. R., & Barnett, B. G. (2014). Considering the Use of Data by School Leaders for Decision Making. In A. J. Bowers, A. R. Shoho, & B. G. Barnett (Eds.), *Using Data in Schools to Inform Leadership and Decision Making* (pp. 1-16). Charlotte, NC: Information Age Publishing.

Bowers, A. J., & Sprott, R. (2012a). Examining the Multiple Trajectories Associated with Dropping Out of High School: A Growth Mixture Model Analysis. *Journal of Educational Research, 105*(3), 176-195. doi:10.1080/00220671.2011.552075

Bowers, A. J., & Sprott, R. (2012b). Why tenth graders fail to finish high school: A dropout typology latent class analysis. *Journal of Education for Students Placed at Risk, 17*(3), 129-148. doi:10.1080/10824669.2012.692071

Bowers, A. J., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal, 96*(2), 77-100. http://muse.jhu.edu/journals/high_school_journal/v096/96.2.bowers.html

Bowers, A. J., Stanton, R., & Boylan, J. F. (2000). *Identification of differentially expressed mRNAs in moderately differentiated human squamous cell carcinomas.* Paper presented at the American Association of Cancer Research annual meeting, San Francisco, CA.

Bowers, A. J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *Journal of Education for Students Placed at Risk (JESPAR), 24*(1), 20-46. doi:10.1080/10824669.2018.1523734

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., . . . Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research, 86*(4), 803-848. doi:10.3102/0034654316672069

Carl, B., Richardson, J. T., Cheng, E., Kim, H., & Meyer, R. H. (2013). Theory and Application of Early Warning Systems for High School and Beyond. *Journal of Education for Students Placed at Risk (JESPAR), 18*(1), 29-49. doi:10.1080/10824669.2013.745374

Carolan, B. V., & Matthews, J. S. (2015). Curricular Choice and Adolescents' Interest in Math: The Roles of Network Diversity and Math Identity. *Teachers College Record, 117*. doi: 10.1177/016146811511701101

Carter, P. L., & Welner, K. G. (2013). *Closing the opportunity gap: What America must do to give every child an even chance*: Oxford University Press.

Chambers, J. M. (1993). Greater or lesser statistics: a choice for future research. *Statistics and Computing, 3*(4), 182-184.

Cleveland, W. S. (1985). *The elements of graphing data*: Wadsworth Publ. Co.

Coleman, C. J. (2021). *Exploring a Generalizable Machine Learned Solution for Early Prediction of Student At-Risk Status.* Columbia University, Retrieved from https://doi.org/10.7916/d8-5scb-n214

Davis, M., Herzog, L., & Legters, N. (2013). Organizing Schools to Address Early Warning Indicators (EWIs): Common Practices and Challenges. *Journal of Education for Students Placed at Risk (JESPAR), 18*(1), 84-100. doi:10.1080/10824669.2013.745210

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline of practice and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics, 26*(4), 745-766. doi:10.1080/10618600.2017.1384734

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, 95*, 14863-14868. Retrieved from http://rana.lbl.gov/papers/Eisen_PNAS_1998.pdf

Färber, I., Günnemann, S., Kriegel, H.-P., Kröger, P., Müller, E., Schubert, E., . . . Zimek, A. (2010). *On using class-labels in evaluation of clusterings.* Paper presented at the MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in conjunction with KDD.

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). *Getting students on track for graduation: Impacts of the Early Warning Intervention and Monitoring System after one year*. Retrieved from Washington, DC: https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=388

Farley-Ripple, E. N., Jennings, A., & Jennings, A. B. (2021). Tools of the trade: a look at educators' use of assessment systems. *School Effectiveness and School Improvement, 32*(1), 96-117. doi:10.1080/09243453.2020.1777171

Farrell, C. C. (2014). Designing School Systems to Encourage Data Use and Instructional Improvement: A Comparison of School Districts and Charter Management Organizations. *Educational Administration Quarterly, 51*(3), 438-471. doi:10.1177/0013161x14539806

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., . . . Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education, 44*(1), 130-160. doi:10.3102/0091732x20903304

Frank, K., Muller, C., Schiller, K. S., Riegle-Crumb, C., Mueller, A. S., Crosnoe, R., & Pearson, J. (2008). The Social Dynamics of Mathematics Coursetaking in High School. *American Journal of Sociology, 113*(6), 1645-1696. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2963869/

Frazelle, S., & Nagel, A. (2015). *A practitioner's guide to implementing early warning systems*. Retrieved from Washington, DC: http://ies.ed.gov/ncee/edlabs/regions/northwest/pdf/REL_2015056.pdf

Freeman, J., & Simonsen, B. (2015). Examining the Impact of Policy and Practice Interventions on High School Dropout and School Completion Rates: A Systematic Review of the Literature. *Review of Educational Research, 85*(2), 205-248. doi:10.3102/0034654314554431

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-prepartory mathematics for students with diverse abilities in early elementary school. *Educational Evaluation and Policy Analysis, 22*(3), 241-254.

Gelman, A., & Vehtari, A. (2021). What are the Most Important Statistical Ideas of the Past 50 Years? *Journal of the American Statistical Association*, 1-11. doi:10.1080/01621459.2021.1938081

Gerzon, N. (2015). Structuring Professional Learning to Develop a Culture of Data Use: Aligning Knowledge From the Field and Research Findings. *Teachers College Record, 117*(4), 1-28. doi:10.1177/016146811511700407

Gleason, P., Crissey, S., Chojnacki, G., Zukiewicz, M., Silva, T., Costelloe, S., & O'Reilly, F. (2019). *Evaluation of Support for Using Student Data to Inform Teachers' Instruction (NCEE 2019-4008)*. Retrieved from Washington, DC: https://eric.ed.gov/?id=ED598641

Grabarek, J., & Kallemeyn, L. M. (2020). Does Teacher Data Use Lead to Improved Student Achievement? A Review of the Empirical Evidence. *Teachers College Record, 122*(12). doi: 10.1177/016146812012201201

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics, 32*(18), 2847-2849. doi:10.1093/bioinformatics/btw313

Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review. *Journal of Youth and Adolescence, 48*(9), 1637-1667. doi:10.1007/s10964-019-01072-5

Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education? In *Equity in discourse for mathematics education* (pp. 17-33): Springer.

Halverson, R. (2010). School formative feedback systems. *Peabody Journal of Education, 85*(2), 130-146. doi: 10.1080/0161956100368527

Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Berkowitz, E., Burnett, T., & Culhane, D. (2020). *A Toolkit for*

Centering Racial Equity Throughout Data Integration. Retrieved from Philadelphia, PA: https://www.aecf.org/resources/a-toolkit-for-centering-racial-equity-within-data-integration/

Heck, R. H., Price, C. L., & Thomas, S. L. (2004). Tracks as Emergent Structures: A Network Analysis of Student Differentiation in a High School. *American Journal of Education, 110*(4), 321-353. doi:10.1086/422789

Hinnant-Crawford, B. N. (2020). *Improvement science in education: A primer*: Meyers Education Press.

Iam-On, N., & Boongoen, T. (2017). Generating descriptive model for student dropout: a review of clustering approach. *Human-centric Computing and Information Sciences, 7*(1), 1. doi:10.1186/s13673-016-0083-0

Ingels, S. J., Pratt, D. J., Alexander, C. P., Jewell, D. M., Lauff, E., Mattox, T. L., . . . Christopher, E. (2014). *Education Longitudinal Study of 2002 (ELS:2002) Third Follow-Up Data File Documentation* (NCES 2014-364). Retrieved from Washington, DC: http://nces.ed.gov/pubs2014/2014364.pdf

Jorion, N., Roberts, J., Bowers, A. J., Tissenbaum, M., Lyons, L., Kuma, V., & Berland, M. (2020). Uncovering Patterns in Constructionist Collaborative Learning Activities via Cluster Analysis of Museum Exhibit Log Files. *Frontline Learning Research, 8*(6), 77-87. doi:10.14786/flr.v8i6.597

Kemple, J. J., Segeritz, M. D., & Stephenson, N. (2013). Building On-Track Indicators for High School Graduation and College Readiness: Evidence from New York City. *Journal of Education for Students Placed at Risk (JESPAR), 18*(1), 7-28. doi:10.1080/10824669.2013.747945

Kim, H., Watkinson, J., Varadan, V., & Anastassiou, D. (2010). Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. *BMC Medical Genomics, 3*(1), 51. doi:10.1186/1755-8794-3-51

Knowles, J. E. (2015). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *Journal of Educational Data Mining, 7*(3), 18-67. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM082

Krumm, A. E., & Bowers, A. J. (2022). Data-intensive improvement: The intersection of data science and improvement science. In D. J. Peurach, J. L. Russell, L. Cohen-Vogel, & W. R. Penuel (Eds.), *The Foundational Handbook on Improvement Research in Education* (pp. 465-484). Lanham, MD: Rowman & Littlefield.

Krumm, A. E., Means, B., & Bienkowski, M. (2018). *Learning Analytics Goes to School: A Collaborative Approach to Improving Education*. New York: Routledge.

Lee, J., Recker, M., Bowers, A. J., & Yuan, M. (2016). *Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data*. Paper presented at the International Conference of Educational Data Mining (EDM), Raleigh, NC. http://www.educationaldatamining.org/EDM2016/proceedings/paper_34.pdf

Lethbridge, T. C. (2021). *Low-Code Is Often High-Code, So We Must Design Low-Code Platforms to Enable Proper Software Engineering*, Cham.

Lucas, C., Wong, P., Klein, J., Castro, T. B. R., Silva, J., Sundaram, M., . . . Yale, I. T. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature, 584*(7821), 463-469. doi:10.1038/s41586-020-2588-y

Mac Iver, M. A., Stein, M. L., Davis, M. H., Balfanz, R. W., & Fox, J. H. (2019). An Efficacy Study of a Ninth-Grade Early Warning Indicator Intervention. *Journal of Research on Educational Effectiveness, 12*(3), 363-390. doi:10.1080/19345747.2019.1615156

Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation, 69*. doi:10.1016/j.stueduc.2020.100842

Moyer-Packenham, P. S., Tucker, S. I., Westenskow, A., & Symanzik, J. (2015). Examining Patterns in Second Graders' Use of Virtual Manipulative Mathematics Apps through Heatmap Analysis. *International Journal of Educational Studies in Mathematics, 2*(2), 1-16. doi:10.17278/ijesim.2015.02.004

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., . . . andMe Research, T. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics, 50*(7), 920-927. doi:10.1038/s41588-018-0151-7

National Academies of Sciences Engineering and Medicine. (2019). *Monitoring educational equity*. Washington, DC: National Academies Press.

National Academies of Sciences Engineering and Medicine. (2020). *Building educational equity indicator systems: A guidebook for states and school districts*: National Academies Press.

National Academy of Education. (2017). *Big Data in Education: Balancing the Benefits of Educational Research and Student Privacy: Workshop Summary*. Retrieved from Washington, DC: https://naeducation.org/wp-content/uploads/2017/05/NAEd-Summary-Report-Big-Data-in-Education-FINAL.pdf

Nitkin, D., Ready, D. D., & Bowers, A. J. (2022). Using Technology to Personalize Middle School Math Instruction: Evidence From a Blended Learning Program in Five Public Schools. *Frontiers in Education, 7*. doi:10.3389/feduc.2022.646471

Oakes, J. (2005). *Keeping track: How schools structure inequality* (Second ed.). New Haven, CT: Yale University Press.

Ólafsson, R. F., & Hansen, B. (2022). Characteristics of the Authority Basis of Icelandic Compulsory School Principals in Comparison to Other TALIS Countries. *Education Sciences, 12*(3), 219. Retrieved from https://www.mdpi.com/2227-7102/12/3/219

Pallas, A. M. (1989). Conceptual and measurement issues in the study of school dropouts. In K. Namboodiri & R. G. Corwin (Eds.), *Research in the sociology of education and* *socialization* (Vol. 8, pp. 87-116). Greenwich, CT: JAI Press Inc.

Pallas, A. M. (1993). Schooling in the course of human lives: The social context of education and the transition to adulthood in industrial society. *Review of Educational Research, 63*(4), 409-447.

Piety, P. J. (2019). Components, Infrastructures, and Capacity: The Quest for the Impact of Actionable Data Use on P–20 Educator Practice. *Review of Research in Education, 43*(1), 394-421. doi:10.3102/0091732x18821116

Piety, P. J., Hickey, D. T., & Bishop, M. (2014). *Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems.* Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics and Knowledge.

Piety, P. J., & Pea, R. D. (2018). Understanding Learning Analytics Across Practices. In D. Niemi, R. D. Pea, & B. Saxberg (Eds.), *Learning Analytics in Education* (Vol. 215-232). Charlotte, NC: Information Age Publishing.

Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-Computer Interaction*: Addison-Wesley Longman Ltd.

R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reardon, S. F. (2019). Educational Opportunity in Early and Middle Childhood: Using Full Population Administrative Data to Study Variation by Place and Age. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 5*(2), 40-68. doi:10.7758/rsf.2019.5.2.03

Reverter, A., Martinez, C., Currey, P., van Bommel, S., & Hudson, N. J. (2020). Unravelling student evaluations of courses and teachers. *Cogent Education, 7*(1), 1771830. doi:10.1080/2331186X.2020.1771830

Romesburg, H. C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.

Rumberger, R. W. (2011). *Dropping Out: Why Students Drop Out of High School and What Can Be Done About It*. Cambridge, Mass.: Harvard University Press.

Sansone, D. (2019). Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics, 81*(2), 456-485. doi:doi:10.1111/obes.12277

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*(7), 912-919. doi:10.1038/s41588-018-0152-6

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies, 10*(1), 30-41. doi:10.1109/tlt.2016.2599522

Selwyn, N., Pangrazio, L., & Cumbo, B. (2021). Attending to data: Exploring the use of attendance data within the datafied school. *Research in Education, 109*(1), 72-89. doi:10.1177/0034523720984200

Sepanik, S., Zhu, P., Shih, M. B., & Commins, N. (2021). *First-year effects of early indicator and intervention systems in Oregon (REL 2021–097)*. Retrieved from Washington, DC: https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=5663

Simon, H. A. (1971). Designing Organizations for an Information-Rich World. In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest* (pp. 37-52). Baltimore, MD: The Johns Hopkins Press.

Soland, J. (2017). Combining Academic, Noncognitive, and College Knowledge Measures to Identify Students Not on Track For College: A Data-Driven Approach. *Research & Practice in Assessment, 12*(Summer 2017), 5-19. Retrieved from http://www.rpajournal.com/dev/wp-content/uploads/2017/07/Summer_2017.pdf#page=5

Stahl, F., Gabrys, B., Gaber, M. M., & Berendsen, M. (2013). An overview of interactive visual data mining techniques for knowledge discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(4), 239-256. doi:10.1002/widm.1093

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics, 33*(1), 1-67. Retrieved from http://www.jstor.org/stable/2237638

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2): Reading, Mass.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-107): Cambridge University Press.

Verplank, W. (2003). Interaction Design Sketchbook: Frameworks for designing interactive products and systems. https://ccrma.stanford.edu/courses/250a/lectures/IDSketchbok.pdf

Wayman, J. C., Shaw, S., & Cho, V. (2017). Longitudinal Effects of Teacher Use of a Computer Data System on Student Achievement. *AERA Open, 3*(1). doi:doi:10.1177/2332858416685534

Weinstein, J. N. (2008). A Postgenomic Visual Icon. *Science, 319*(5871), 1772-1773. doi:10.1126/science.1151888

Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician, 63*(2), 179-184. doi:10.1198/tas.2009.0033

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . the Major Depressive Disorder Working Group of the Psychiatric Genomics, C. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics, 50*(5), 668-681. doi:10.1038/s41588-018-0090-3

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science, 2*(2), 165-193. doi:10.1007/s40745-015-0040-1

Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. *IEEE Trans On Neural Networks, 16*. doi:10.1109/tnn.2005.845141

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences, 45*, 1-37. doi:10.1017/s0140525x20001685

Zhao, Y., & Karypis, G. (2005). Data clustering in life sciences. *Molecular Biotechnology, 31*(1), 55-80. doi:10.1385/MB:31:1:055

Zimmermann, A. (2020). Method evaluation, parameterization, and result validation in unsupervised data mining: A critical survey. *WIREs Data Mining and Knowledge Discovery, 10*(2), e1330. doi:https://doi.org/10.1002/widm.1330

Appendix A: Column order (left to right) of high school and post-secondary course grade columns in Figures 2 and 3.

High School Courses by CSSC				Post-Secondary College & University by CCM		
Order	**CSSC Label**	**Description**		**Order**	**CCM Label**	**Description**
1	LETTERS	Letters/English		1	ENGLISH	English language and literature/letters
2	MATH	Mathematics		2	SOCIALSC	Social sciences
3	PUBLAFF	Public affairs		3	MATH&SCI	Mathematics and statistics
4	CITIZEN	Citizenship/civic activities		4	VISUART	Visual and performing arts
5	FORLANG	Foreign languages		5	BIOSCIEN	Biological and biomedical sciences
6	TRANSPO	Transportation and material moving		6	PSYCHOLO	Psychology
7	THEOLOG	Theology		7	BUS&MARK	Business/management/marketing/related
8	LIFESCI	Life sciences		8	PHYSICAL	Physical sciences
9	INTEPER	Interpersonal skills		9	HISTORY	History
10	VOCAECO	Vocational home economics		10	COMMUNIC	Communication, journalism, related
11	PEAWARE	Personal awareness		11	RELIGSTU	Philosophy and religious studies
12	COMPUTER	Computer and information sciences		12	FORELANG	Foreign languages/literature/linguistics
13	PARK&RE	Parks and recreation		13	HEALTHSC	Health/related clinical sciences
14	BUS&OFF	Business and office		14	COMPUTER	Computer/information science/support
15	MILITEC	Military technologies		15	PARKSSTU	Parks/recreation/leisure/fitness studies
16	BASICSK	Basic skills		16	EDUCAT	Education
17	MECHANI	Mechanics and repairers		17	ETHNIC	Area/ethnic/cultural/gender studies
18	SCIENCE	Science technologies		18	LIBERART	Liberal arts/sci/gen studies/humanities
19	COMMUNI	Communications		19	FAMILSER	Family/consumer sciences/human sciences
20	ETHINIC	Area and ethnic studies		20	MULTISTU	Multi/interdisciplinary studies
21	INDARTS	Industrial arts		21	BASICSKI	Basic skills/remedial education
22	BUS&MAN	Business and management		22	OTHER	Other
23	SPECCUR	Special education-resource curriculum		23	SECURITY	Security and protective services
24	MILISCI	Military sciences		24	ENGINEER	Engineering
25	CONSTRU	Carpentry		25	HEALTHSK	Health-related knowledge and skills
26	MAR&DIS	Marketing and distribution		26	ENGITECH	Engineering technologies/technicians
27	COMTECH	Communication and technologies		27	PUBSERVI	Public administration/social service
28	HEALTH	Allied health		28	NATURAL	Natural resources and conservation
29	LIB&GEN	Liberal/general studies		29	LEISURE	Leisure and recreational activities
30	CAREER	Special education-vocational, career preparation, career exploration		30	LEGALSTU	Legal professions and studies
31	AGRIBUS	Agribusiness and agricultural production		31	THEOLOGY	Theology and religious vocations
32	ENGITEC	Engineering and engineering related technologies		32	AGRICULU	Agriculture/operations/related sciences
33	LAW	Law		33	SOCIALSK	Interpersonal and social skills
34	HEALACT	Health related activities		34	MECHANIC	Mechanic/repair technologies/technicians
35	SOCISCI	Social sciences		35	ARCHITEC	Architecture and related services
36	EXECINT	Executive internship		36	COMMUTEC	Communication technology and support
37	AGRISCI	Agriculture sciences		37	PERSONAL	Personal and culinary services
38	SPECEDU	Special education		38	CONSTRUC	Construction trades
39	PHILOSO	Philosophy and religion		39	PERSAWAR	Personal awareness/self-improvement
40	LEISURE	Leisure and recreational activities		40	RESEROFF	Reserve officer training (JROTC, ROTC)
41	PSYCHOL	Psychology		41	PRECIPRO	Precision production
42	CONSUME	Consumer, personal, and miscellaneous service		42	CITIZEN	Citizenship activities
43	ENGINER	Engineering		43	TRANSPOR	Transportation and materials moving
44	NATURAL	Renewable natural resources		44	RESIDENPR	Residency programs
45	LIBR&AR	Liberal and archival sciences		45	LIBRASCI	Library science
46	ARCHITE	Architecture and environmental design		46	MILITECH	Military technologies
47	PRECPRO	Precision production		47	SCITECH	Science technologies/technicians
48	HEALSCI	Health sciences		48	HSCERTIF	HS/Secondary Diplomas/Certificates
49	VIS&PER	Visual and performing arts				
50	PHYSICA	Physical sciences				
51	EDUCATE	Education				
52	PROTSER	Protective services				
53	MULTIDI	Multi/interdisciplinary studies				
54	HOMEECO	Home economics				