Educational Attainment: An Agent-Based Model

by

Anna Truman

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in the

Mathematics

Program

YOUNGSTOWN STATE UNIVERSITY

May, 2022

Educational Attainment: An Agent-Based Model

Anna Truman

Signature:

Anna Truman, Student — Date

Approvals:

Dr. Alicia Prieto-Langarica, Thesis Advisor — Date

Dr. Alexis Byers, Committee Member — Date

Dr. Lucy Kerns, Committee Member — Date

Dr. Alejandra Donají Herrera Reyes, Committee Member — Date

Dr. Salvatore A. Sanders, Dean of Graduate Studies — Date

Abstract

Educational attainment is a subject of great importance in today's world. The challenge to maximize educational outcomes is now seen on a national scale. There have been many studies on the interventions and factors that can be used to improve educational attainment. Using interventions that have proven effective in these studies, an original survey exploring the impact of these interventions on educational attainment is collected and combined with census data. It is then integrated into an agent-based model that predicts educational outcomes. Controlling for different variables, this agent-based model gives insights into the influence of different intervention combinations.

Acknowledgements

There are many people who have put in time, effort, and support to get me to where I am now. I would first like to thank Dr. Alicia Prieto-Langarica, my thesis advisor, for taking me under her wing and encouraging me to find something that I am passionate about. It is easy to fall into the trap of needing to constantly achieve, but Dr. P has the gift of seeing beyond this. She taught me the importance of being interested and excited in what you do, a lesson that she lives every day in her unending ability to inspire.

I would also like to thank Dr. Alejandra Donaji Herrera Reyes for all of the time she spent answering my endless questions. Though she is on a different continent, she was always a zoom call away. Those weekly zoom meetings were always a bright spot in my day.

Additionally, I would like to thank the Youngstown State Mathematics and Statistics Department. The professors here are kind, helpful, and brilliant. It has been an honor working with such knowledgeable people.

Another thank you goes to my undergraduate professor, Dr. Gary Thompson of Grove City College. Without his encouragement and love for mathematics, I would never have had the courage to apply to graduate school. He was a true inspiration during my undergraduate years, and I am so grateful for all of the difficult tests, confusing homework, and Putnam problems he challenged me with.

Finally, I would like to thank my family and friends. It has been a long journey from start to finish, and I have been privileged with their constant support. Truly the greatest honor in life is to have such wonderful people surrounding you, and I am blessed beyond comparison.

Table Of Contents

Introduction

Educational attainment is highly linked to the perception of success, wealth, and stability. There are many people who believe that a higher education will lead to higher quality of life. While this is not always true, it is undeniable that higher education is a major factor in the lives of Americans. One problem with the current educational system is that not all students are given the same academic opportunities. There are wide disparities in the quality of public education based on students' school districts. In addition, different factors – such as finances, lack of encouragement, lack of exposure to educational options, etc. – can have great impact on a students' ability to succeed academically. The aim of this research is to determine some interventions that will positively affect students and give them a strong and more equal start in their future endeavors.

In this study, the researchers constructed an agent-based model for educational attainment. After gathering data from the United States Census Bureau and conducting their own primary survey, the researchers were able to build this model. Considering factors of gender, race, interventions, and parental education, the model aims to give a relatively comprehensive view of general factors that influence students. Using probability theory to calculate the expected impact of these factors, the model is able to take in a person with predefined attributes and predict their educational attainment outcome. This model can then be used to compare educational outcomes between interventions.

1.1 Ohio Demographics

Unsurprisingly, the Census Bureau harbors a wealth of data regarding the educational attainment and demographic landscapes of Ohio. According to an article from 2021 , 90.4% of Ohioans 25 or older had received a high school diploma or GED, and 28.3% had received a Bachelor's degree or higher in the years 2015-2019 [12]. These statistics are slightly different from the national averages of 2015, which showed that 88% of adults had at least a high school diploma or GED while 33% adults had a bachelor's or higher degree [7]. With the updated 2020 census, that national average degree percentage has increased to 37.5% of adults 25 + having a bachelor's degree or higher. This proportion has been steadily increasing since 1940 [10]. With such an impact on America's society and economy, the importance of educational attainment will only continue growing along with the proportion of people achieving higher degrees.

The data available from the Census Bureau is primarily demographic data. Though this is not surprising, such data can only reach so far when studying the complex topic of educational attainment. The demographic spread of Ohio from the 2019 data is found in the following table.

Race	%	**Sex**	%
White	78.4	Female	51
African American	13.1	Male	49
American Indian	0.3		
Asian	2.5		
Hispanic or Latino	4.0		
Other/NA	1.7		

Table 1.1: 2019 Ohio Demographics taken from Census Bureau of Statistics [12]

Though the demographics will likely have an impact on educational attainment, the data begs the question: what other factors could be important or impactful than demographic data? This question sparked the project and led to many different considerations. The first field the researchers wanted to address was parental education attainment, i.e. how does a parent's educational attainment level affect that of their child. Finding the data for such a question presents a challenge in itself, though there are a few studies that have considered this field.

Literature and Interventions

2.1 Introduction

The first step of conducting such research lies in defining the interventions to study. When considering different options, the researchers wanted to focus on interventions that have been successfully used in some form of schooling, and also interventions that are reasonable to employ. It was important to consider the cost and effort of each intervention because these interventions need to be possible at every public school, regardless of size and funding. For example, while a mandatory course in how to plan for the future and take next steps would certainly be helpful, it is unrealistic to implement state-wide in public schools. Thus, there must be a good middle ground of plausible implementation.

2.2 The Factor of Parental Educational Attainment

A 2009 article written by Dubow, Boxer and Huesmann, used data from a 40-year longitudinal study[2]. In this article, the authors "examine the prediction of individuals' educational and occupational success at age 48 from contextual and personal variables assessed during their middle childhood and late adolescence'.'

This article took data from the Columbia County Longitudinal Study. The study ranged from 1960 to 2008, starting with 856 third graders who were interviewed along with their parents. The study followed up with subjects at ages 19, 30, and 48. Using multiple factors from the data, the authors built a predictive cognitive-ecological model to predict educational attainment and occupational attainment at age 48. Some of the specific factors considered were parental education, parental occupation, IQ, child aggressiveness, adolescent aspirations, socioeconomic status, and negative family interaction during childhood (parental rejection, corporal punishment, and parental disharmony).

Researchers controlled for household income, family models, and other demographic factors. With their model, they discovered that "parental educational level during childhood had no direct effects on educational level at age 48 but had significant indirect effects mediated through age 19 educational aspirations and age 19 educational level. Thus, children with more highly educated parents developed higher aspirations for their own education and attained more education by age 19, which in turn related to higher levels of adult educational attainment" [2]. Furthermore, of all the factors studied, parental education and subject IQ had the largest positive effects on the subjects' educational attainment outcomes. These results lend credence to the decision to include parental educational attainment in the model for this thesis.

Notably, the longitudinal study does not simply report the factors that influence attainment. The paper also looks at the reasoning behind each influence. While subject IQ was deemed a direct influence on educational attainment, parental education was found to be indirect, mostly linked to a child's educational aspirations, not their abilities. A positive relationship between parent and child education exists likely because parents with higher education levels encourage achievement and offer support when figuring out the steps to higher education. So, when considering interventions for the study in this paper, encouraging aspirations and providing support may prove to be effective.

2.3 Recommended Interventions from Supported Studies

Some other interventions are discussed in Belfield and Levin's 2007 article [1]. A similar discussion can be found in Ou's longitudinal study [6]. Both of these papers compare different intervention studies to see which interventions appear to be most effective. The first paper finds that reduced classroom size "demonstrably raises the high school graduation rate" [1]. Other factors in that same study include supportive teachers, high parental involvement, and instructional improvement efforts. From these possible interventions, reduced classroom size and level of parental involvement are unfortunately too difficult to change or control when it comes to public schools. However, supportive teachers is a possible intervention that could be employed.

The second paper explores some of the same interventions, but it adds that school support is an intervention that makes a statistically significant difference in educational attainment outcomes. School support is defined in the following passage.

> Measures of school quality and school mobility were defined as school support because they were viewed as potential sources of influence on student educational attainment. School quality was defined as any attendance in a magnet school from grades

> 4 through 8, and an assessment of school characteristics in fifth grade... The other indicator, school characteristics, was measured as the average percentage of low-income families, mobility, and truancy at the school level in fifth grade... School mobility was measured by the number of times participants changed schools between grades four and eight. School mobility was found to predict educational attainment and mediated the effect of an early intervention program (Temple and Reynolds, 1999). Thus it was used as an indicator of the school support hypothesis. [6]

Though the school support intervention is mostly focused on magnet schools, there are qualities of magnet schools that can be used as interventions at a typical public school. Magnet Schools of America explains that magnet schools are a full-service version of the typical public school. They focus on individually themed curricula, as well as soft skills that give a more holistic learning experience to students [4]. A key component to many magnet schools is free transportation, which allows for more socioeconomic equality across the student population.

This transportation aspect reminded the author of her own public high school, where free transportation was used as an intervention to encourage higher education. In this experience, the faculty organized mass college visits to four different colleges. The colleges were all of different size and type (small private, large private, state, and specialty) and each student was allowed to sign up for one school. During a school day, the entire junior class boarded buses and went to their respective college visits. This intervention is as attainable as any other field trip, and it allows students who may have never had the chance to visit a college to experience that. Most college visits are scheduled during the work week and normal working hours, so having a guardian that is able to take their child on a college visit is a privilege that is not available to everyone. Thus by having a school-funded college visit, every student is given an equal opportunity. Because of this, the researchers deemed this a worthy intervention to consider.

Another idea that can be gained from magnet schools is specialized curriculum. Though students have to meet requirements to get their high school diploma, they are still able to tailor their schedules in a way that is of more interest to them. Probably the highest level of specialized education in public schools comes in the form of college preparatory courses, which actually allow students to experience what they may study in the future. These courses can also count towards college credit at some institutions, lessening the amount of money students would have to spend on tuition in the future. Furthermore, public schools can offer academic clubs that are able to go beyond the scope of the curriculum. These clubs can range from speech and debate club to robotics club to agricultural club. Giving students the freedom to think outside of classroom restrictions gives them the opportunity to get a more realistic view of their interests as well as letting them interact with others in the community of their chosen field.

Beyond these tactical interventions lies a more abstract support system to emulate. Magnet schools focus not only on the academics, but also on community and team building.

> Through a more harmonious and healthy interaction to various cultures and socioeconomic backgrounds, to developing a deeper understanding of community that comes from hands-on interaction with corporations, non-profits, cultural and academic institutions, students are exposed to a microcosm of the world at large, learning skills of interaction, team building and cooperation. [4]

Skills that come with real life experience are difficult to parallel in a school environment, but maximizing exposure to new situations is a good start. School sports and extra-curricular activities offer diverse interactions, team building, and experience in dealing with conflicting opinions.

Magnet schools provide great insight into lesser known interventions that might be applicable to public schools. However, this project is not arguing for turning public schools into magnet schools. Magnet schools tend to be extremely specialized in their curriculum which is not always the best approach for student educational attainment. Also, as they are known for small class sizes, it would take multiple magnet schools to take on the student population of a single public school. As far as the American educational infrastructure goes, it makes much more sense to tweak interventions in a public school than to convert public schools to magnet schools. The public school system in America can provide great opportunities for students, and this research aims to find out how to maximize those opportunities, as well as equalizing them across the student population.

2.4 Interventions in the Model

The interventions considered for the model are as follows: college preparatory courses, college visits organized by the high school, an inspiring teacher/mentor, faculty encouragement, school sports, academic clubs, and extra-curricular activities. Many schools already have college preparatory courses, sports, clubs, and extra-curricular activities. However, they can use these as interventions by intentionally encouraging more students to participate and by offering transportation to and from these activities. Faculty encouragement and inspiring teacher/mentor are more abstract interventions to facilitate. Schools can “implement” this by simply raising the awareness of the impact teachers make. In the survey data collected for this research, the intervention that had the highest impact on educational attainment was an inspiring teacher/mentor (data to be discussed in Section 3.3). Sometimes the importance of educators can get lost in the requirements and objectives of academia. But the fact remains that teachers are more impactful than they realize, and encouraging educators is an excellent way to help students [1]. Possibly the most novel intervention comes in the form of college visits organized by the high school. This is certainly a sensible intervention, as most public schools offer field trips for students - schools could set up field trips to nearby colleges during a school day. It also is the least common intervention, as will be seen with the survey data (Section 3.3). All of these interventions are accessible at most public schools, and those that are not could be reasonably instituted. The importance of parental educational attainment will also be recorded as an added demographic to study.

2.5 Demographics in the Model

As stated before, parental educational attainment is included in the model as one of the demographic variables. Other than that, race and gender are also included. Parental educational attainment is divided into first generation (neither parent achieved a degree past high school) and not first generation (one or both parents achieved a degree past high school). The values for gender are male and female. The researchers decided to limit race to three values: white, black, and other. This was decided to reflect the census data in Table 3.1. The researchers strongly considered including Asian as a separate value for race, but due to the American perception of race, there is generally different academic bias towards those with a western Asian background than an eastern Asian background. Yet this bias is not recorded in the available census data, so any possible advantage that those who identify themselves as eastern Asian may have is lost among the biases for those of western Asian identification. The same reasoning applied to the other races recorded in the census data. Though the individual experiences for each race are unique, academic biases tend to be similar regardless.

Data

3.1 Data Collection

This research combined preexisting data with original survey data. The preexisting data was obtained from the United States Census Bureau [11]. Due to the recent upheaval in education, the census data is from 2018 to ensure that the baselines are not impacted by Coronavirus. The original data consisted of a survey. Starting in April 2021, the Youngstown State Institutional Review Board (IRB) accepted this survey proposal with protocol number 090-21 to be viable to distribute to the general public. The only requirement for participation was that subjects had to be 18 years or older. Using multiple different social platforms, the researchers spread the survey, trying to target diverse groups of Ohioans.

Due to the restrictions set by the IRB, the survey was completely voluntary and offered no rewards for participation. As a result, survey responses were fairly sparse, and mainly made up of women who identified themselves as white. According to an article written by Saleh and Bista, this population is to be expected, as women generally are more likely to answer online surveys [8]. They also found that people are much more invested in surveys that have subjects of interest to them. Furthermore, their research discovered that, "... participants prefer completing electronic surveys received mostly from students, colleagues and authority figures (e.g. department chair or higher) compared to people from other organizations who they do not know personally or professionally." Since this survey was first distributed to as broad of an Ohio population as possible, most people who received it were strangers who likely did not have a vested interest in the research.

After peaking at about 35 responses from months of posting, the researchers decided to send the survey to colleagues and acquaintances in order to garner more responses. This proved fairly successful, though it did skew the data even further towards white women as well as people with above average educational attainment when compared to the census data. Due to Coronavirus, community gatherings and other group settings were not common, so the researchers were unable to gather many responses that way. The few community gatherings they did attend added some responses, but still not enough to make a decent sample size. Though not as much as the researchers hoped for, the survey amassed 138 responses by the time data analysis rolled around, with 128 of those responses without any missing data.

These imperfections in the data are to be expected and are an excellent learning experience for future ventures. In undergraduate and graduate classes, data is typically cleaned and formatted for the students to use. However, the most difficult part of data analysis in the real world is collecting the data and transforming it into usable information. By experiencing this first-hand, the author will be able to have realistic expectations when conducting similar research in her future career, as well as learning some helpful tactics to increase response rates when such a challenge arises.

3.2 Census Data

The census data comes in 3 different Excel spreadsheets: all races, white alone, and black alone. Each spreadsheet is a cross tabulation of age and sex vs. educational attainment levels. The population this data was taken from are people over the age of 18 from the United States. These detailed tables were not available on a state-by-state basis so the population is national and not limited to Ohio. Because those who are in their early twenties have not had the opportunity to achieve higher education, this study focused on the data from those 25+.

With race broken down into white, black and other in the model, the author calculated the population of other by subtracting the white and black values from the all races values. In total, there were 172,685,000 subjects in the white 25+ population, 27,047,000 subjects in the black 25+ population, and 20,098,000 subjects in the other 25+ population.

A cross-tabulation of the data can be seen in Figure 3.1 below. The data is broken down by gender and race, with the percentages representing the proportion of that subset that completed each level of education. Note that to complete middle school, a subject must first complete elementary school. Thus there is a reverse-cumulative pattern in each row when moving from left to right. This data does assume that those who have completed a bachelor's degree have also completed an associate's degree. Though this is not true in real life, it is true that people who complete a bachelor's degree have earned more credits than are required for an associate's degree. This line of thinking allows for the educational attainment variable to be ordinal. The same follows for a Doctorate and a Master's degree. Though it is possible to earn a doctorate without a Master's degree, a person with a Doctorate has completed more schooling than somebody with a Master's. Additionally, by viewing the levels of schooling as ordinal, the researchers were able to eliminate the need for a "years of schooling" variable which is much more subjective than a degree. While one person could complete their Bachelor's in 3 years, another person could take 6 years to complete it. Thus, completion of a degree is much more useful than years of schooling.

Sex	Race	Total Count (in thousands)	Percent of subset that completed level of education							
			None	Elementary	Middle	High	Associate's	Bachelor's	Master's	PhD
Male	White	84276	100.00%	99.10%	97.58%	89.64%	39.87%	34.90%	12.67%	2.39%
	Black	12196	100.00%	99.29%	98.59%	87.73%	28.30%	23.17%	7.54%	1.14%
	Other	9390	100.00%	99.94%	99.77%	85.15%	16.96%	14.50%	2.41%	0.53%
Female	White	88409	100.00%	99.09%	97.75%	90.79%	42.24%	35.53%	13.24%	1.56%
	Black	14852	100.00%	99.19%	98.52%	88.12%	33.77%	26.91%	10.01%	1.26%
	Other	10708	100.00%	99.46%	99.10%	85.59%	22.30%	18.87%	2.53%	0.06%

Figure 3.1: Breakdown of Census Data Percentages

3.3 Survey Data

As stated above, the survey data contains 138 responses. Only 128 of these responses were complete enough to include in analysis. With such a small response, the subsets of data were very small. However, following the same process with more responses would lead to more reliable outputs.

The survey collected the variables seen in the census data (race, sex, and educational attainment outcome) to show how the survey data compares to the much more extensive census data. See Figure 3.2 to for the breakdown of these variables.

Sex	Race	Total Count	Percent of subset that completed level of education							
			None	Elementary	Middle	High	Associate's	Bachelor's	Master's	PhD
Male	White	42	100.00%	100.00%	100.00%	100.00%	59.52%	52.38%	28.57%	7.14%
	Black	1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	Other	3	100.00%	100.00%	100.00%	100.00%	66.67%	33.33%	0.00%	0.00%
Female	White	80	100.00%	100.00%	100.00%	98.75%	70.00%	61.25%	33.75%	6.25%
	Black	1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%	0.00%
	Other	1	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%	0.00%

Figure 3.2: Breakdown of Survey Data Percentages

Perhaps most notably, the survey data had all but one individual earning a high school degree. This is very different from the percentages seen in Figure 3.1, where not a single subset had 100% attainment for even elementary school. Furthermore, this data is only viable for the subset of white female, and perhaps white male. The black and other categories had very few responses. Even amongst white males and females the survey data seems to have captured highly skewed educational attainment outcomes, with the expected doctorate percentages about 6 times higher in the survey data than the census data. Yet the importance of the survey data was to record other variables that were not available in the census data. Such variables that are collected in the survey were not available in any data sets found in the initial research on the topic. Information on interventions, as well as parental educational attainment, were generally part small longitudinal studies that did not offer the type of data needed for this thesis.

Beyond the three variables in Figure 3.2, the survey collected a wide range of variables. A list of all of the questions is included in Appendix A. With this survey, the researchers were able to to collect information on the interventions discussed in previous sections. Some of the questions were open-ended, such as "What was the most positive impact on your academic career?" The responses to such questions were not used in the building of this agent-based model. For further research with this data they may provide great insight. However, this paper will only explore the categorical data that was recorded in questions 3-9 and 14-18.

Due to the complexity of the agent-based model, the researchers decided to include only two of the most promising interventions found in the data. The reason for such complexity will be discussed later in the paper, but two was found to be the best number. Studying the survey responses gave some interventions a stronger interest than others.

Figures 3.3, 3.3, and 3.3 provide a good visualization of what the data looked like. If a subject selected "Yes" to the question in the title, they were then asked to rate the impact of that intervention on their decision to pursue higher education. The three levels of impact were defined as:

1. Minimal Impact: This experience had nothing to do with my decision to pursue higher education. I would have made the same decision regardless of this experience.

2. Medium Impact: This experience definitely strengthened my decision to pursue higher education. Without this experience, my decision may have been different.

3. High Impact: This experience was one of the sole reasons for my decision to pursue higher education. My decision would have been different without this experience.

These levels of impact allowed the researchers to determine which interventions seemed most promising to include in the model. As can be seen in those three figures, a good proportion of individuals answered "Yes" to each question. For the question on academic clubs, it appears that a majority of affirmative responses did not find much impact in that activity. Because of this, the researchers decided not to include academic club in the model for this paper. The question of faculty encouragement had a better spread of impact amongst the affirmative answers, with most people saying there was a medium impact on their decision to pursue higher education. However, this question was very similar to that of Figure 3.3, so ultimately the researchers decided not to include faculty encouragement. Finally, Figure 3.3 investigated the impact of college preparatory courses. Once again, this had a great affirmative response rate that had a pretty even distribution of impact levels. For this intervention, it was difficult to decide whether or not it should be included in the model. In the end, there were other interventions that afforded more promise for the model, so college preparatory classes were not included.

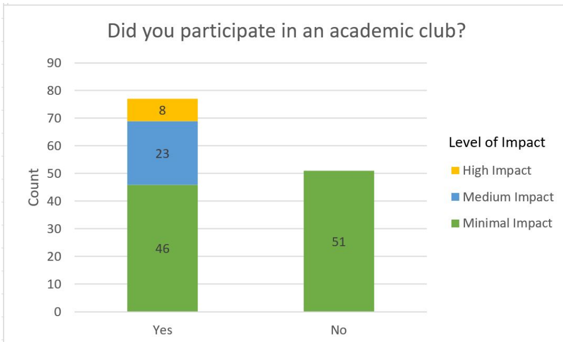

Figure 3.3: Survey results for the academic club intervention.

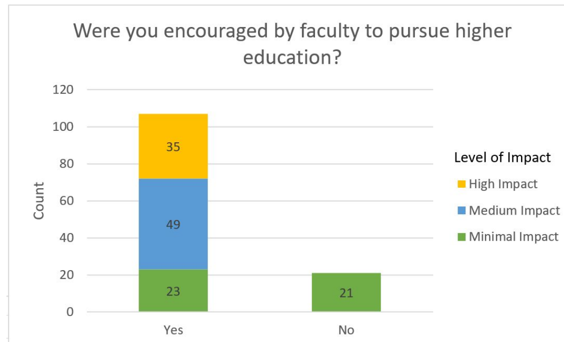

Figure 3.4: Survey results for the faculty encouragement intervention.

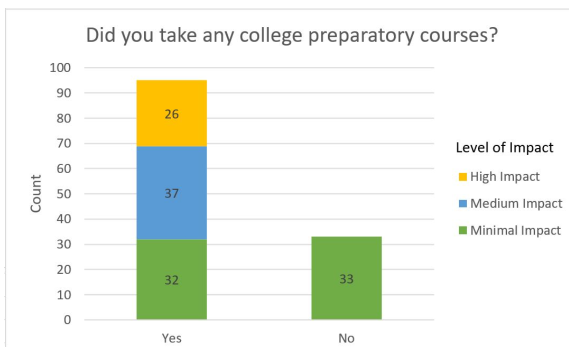

Figure 3.5: Survey results for the college preparatory courses intervention.

In Figure 3.3, the question explores the impact of an inspiring teacher or mentor. Note that the distribution of "Yes" to "No" is very similar to that of Figure 3.3. However, with an inspiring teacher/mentor, more than 50% of people said that it had a high impact on their decision to pursue higher education. Thus, the researchers chose to include this as an intervention in their model over that of an encouraging faculty member.

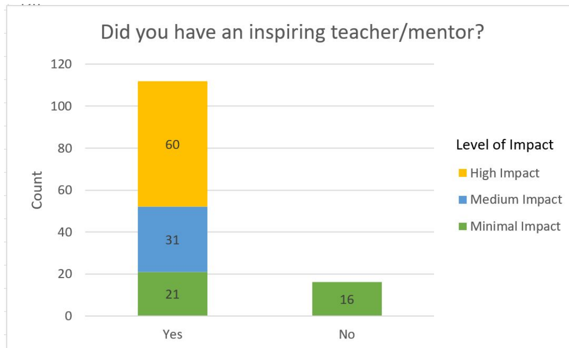

Figure 3.6: Survey results for the inspirational teacher intervention.

The last intervention that had interesting data was school sponsored college visits. As can be seen from Figure 3.3, only 32 of the 128 people had this intervention, meaning that it is the intervention that has the biggest opportunity for fresh implementation across districts. Since this research is focused on finding interventions that can make a difference, choosing a lesser known intervention and seeing if it can positively impact educational outcomes is the best option to include in the model.

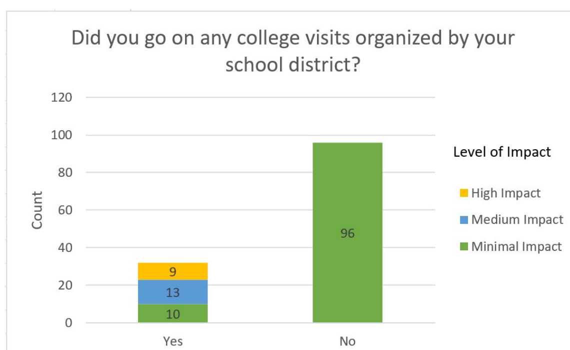

Figure 3.7: Survey results for the school-funded college visit intervention.

Methods

4.1 Agent-based Models

Agent-based models (ABMs) are a newer field in the world of mathematical modeling. The term itself explains the setup and the purpose of the model - the model is made of autonomous *agents* that interact with each other based on a set of rules. With these rules, complex dynamics can be modelled between the agents, as well as each agent's interaction to its environment. The researchers used MATLAB as the coding environment in this project. In a book by Macal and North, the authors give an excellent introduction to the power and reasoning behind ABM.

> Agent-based modelling and simulation (ABMS) is a relatively new approach to modelling complex systems composed of interacting, autonomous 'agents'. Agents have behaviours, often described by simple rules, and interactions with other agents, which in turn influence their behaviours. By modelling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviours can be observed as it gives rise to the behaviour of the system as a whole. By modelling systems from the 'ground up'– agent-by-agent and interaction-by-interaction—self-organization can often be observed in such models. Patterns, structures, and behaviours emerge that were not explicitly programmed into the models, but arise through the agent interactions. The emphasis on modelling the heterogeneity of agents across a population and the emergence of self-organization are two of the distinguishing features of agent-based simulation as compared to other simulation techniques such as discrete-event simulation and system dynamics. Agent-based modelling offers a way to model social systems that are composed of agents who interact with and influence each other, learn from their experiences, and adapt their behaviours so they are better suited to their environment. [3]

In the case of this research, the current version of the model looks at an agent's reaction to its environment rather than its relationships with other agents. Each agent in the model represents a specific person with given attributes. This model will show how different interventions and circumstances affect the agent; the goal is to find the most effective interventions for different groups so that educational attainment outcomes are maximized.

The flexibility of an ABM makes it an excellent choice for this type of research. Due to the challenges of trying to collect data, the output of the model is not as reliable as the researchers wanted. However, with updated data, the model can retain the same structure and produce better outcomes. Hence, the agent-based model is dependable, but the output's strength is determined by the available data. Additionally, any influx of new variables can be added in with relative ease.

4.2 Motivation

The most important question of all is: why is this model important, and how can this model be useful? Agent-based modeling is a unique type of model that can easily grow and change with new information. While running a multiple regression model would answer the same questions that are discussed in the Results section, the agent-based model can be tweaked and adjusted to fit any new influx of information. Because of this flexibility, the research conducted on this project does not need to stop with this paper. More students and more minds will be able to take the current work and go farther than any single graduate student can in their time at Youngstown State University. The ultimate goal of this research is to expand the current ABM to include agent-to-agent interaction and eventually build a model that can accurately reflect a real high school. More expansions of the research will be discussed in the Future Works section below. Furthermore, the subject of this research is especially important for the upcoming generation of students. Due to Coronavirus, middle school and high school students were subjected to at least one year of online or hybrid learning, with some still not fully back in person. As stated in the Literature and Interventions section, the social aspect is crucial to educational attainment. While schools start to return to in-person classes, social and academic interventions will be vital to convince these students to put forward the effort to catch up and carry on in this new era of learning.

4.3 Probability and Population

The goal of this project was to form a model that can input a person and output their expected educational attainment outcome based on different characteristics. As discussed above, the demographic characteristics that are included in the model are gender, race, and parental educational attainment (whether or not that person is first generation). The interventions that are included are school sponsored college visits and inspirational teachers/mentors. Thus, each person is given 5 different attributes in total, labeled as: gender, race, cvis, inspteach, firstgen. The former two attributes are encoded as 0 for “male” and 1 for “female”; 0 for “white”, 1 for “black”, and 2 for “other”. Both inspteach and cvis are binary variables, encoded 0 for “no” and 1 for “yes.” The firstgen variable is encoded as 0 for “first generation” and 1 for “not first generation”. With this set of attributes, there are 48 possible combinations. In the MATLAB code, each combination is assigned a vector based on the attribute encodings. For example, a female other with no college visit, an inspirational teacher, and a parent who got their Master’s degree would be represented by the vector [1 2 0 1 1]. For each person, there are 8 levels of education that they can achieve: none, elementary school, middle school, high school, associate degree, bachelor degree, master’s degree, and doctorate.

In order to predict what the outcome would be, the conditional probability for that person has to be calculated. According to Ott and Longnecker, the conditional probability is defined as follows.

Definition 4.3.1. [5] *Consider two events A and B with nonzero probabilities, $P(A)$ and $P(B)$. The conditional probability of an event A, given event B, is*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

The conditional probability of B, given event A, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Of course, this definition can have different outcomes based on the dependence of the events.

Definition 4.3.2. [5] *Two events A and B are independent events if $P(A|B) = P(A)$ or $P(B|A) = P(B)$.*

For this data set, the hope is that the events are not independent. If the events are not independent, that means the interventions and demographic variables do have an impact on the expected educational outcomes.

When calculating these conditional probabilities, the greatest struggle is to have enough data.

For example, say there is a person who is encoded as $[0\ 0\ 0\ 1\ 1]$. Then that person is a white male who did not go on any school-sponsored college visits, did have an inspirational teacher, and is not first generation. Then the probability of him getting a high school degree would be found as follows:

$$P(high\ school \mid male,\ white,\ no\_cvis,\ yes\_inspteach,\ yes\_firstgen) = \frac{P(high\ school\ \cap\ male,\ white,\ no\_cvis,\ yes\_inspteach,\ yes\_firstgen)}{P(male,\ white,\ no\_cvis,\ yes\_inspteach,\ yes\_firstgen)}.$$

Were the data evenly distributed between all 48 combinations, the numerator would be calculated from only $\frac{1}{48}$ of the data. With a survey data set of only 128, this would be nearly impossible.

However, the data was extremely biased towards white females, with 80 responses falling into that category. Though 80 is not a massive number, it is enough to build a rudimentary model that displays outcomes based on the data that was collected.

Thus, there are two separate models that were built for this project. The first model, referred to as the **White Female Model**, takes the collection of white female responses as the population for the data. Due to the distribution of responses, this model leaves out the cvis variable, leaving only four possible combinations of people in the white female population:

1. Did not have an inspirational teacher and is first generation.

2. Did have an inspirational teacher and is first generation.

3. Did not have an inspirational teacher and is not first generation.

4. Did have an inspirational teacher and is not first generation.

The second model, denoted the **General Model**, still considers the 48 possible combinations. However, it uses data that is generated from a mix of the census data and the survey data. With this generated data as the population, the conditional probabilities can be calculated for each level of schooling and each combination of person. Though the implications seen from this model are not based on real life data, it shows the capabilities of what the model could do given enough survey data.

4.4 MATLAB Code Structure

The code created for the ABM has the general structure seen in Figure 4.8. While the White Female Model has externally calculated conditional probabilities, the General Model requires more work.

The full extent of the White Female Model can be seen in Figure 4.8: creating a matrix of people (where each row represents one person), running that matrix through an Educ_Out function that takes into account the conditional probabilities, and outputting the educational outcome number. That outcome number is then added to a new matrix, called Outcome. There is additional code that outputs a bar graph which displays the counts of different outcomes. In order to label the exact count for each outcome, a function designed by Elimelech Schreiber, called barvalues, was employed [9]. It is important to note that in the White Female Model, there are only 4 cases in the Educ_Out function, as opposed to the 48 cases seen in the General Model.

The General Model uses the same Educ_Out function, however it requires much more code to create data and calculate the conditional probabilities for the 48 cases. The details of this data generation are discussed in Section 4.5. Once the data is generated, the conditional probabilities are calculated and entered into Educ_Out function, and the steps seen in Figure 4.8 lead to the output.

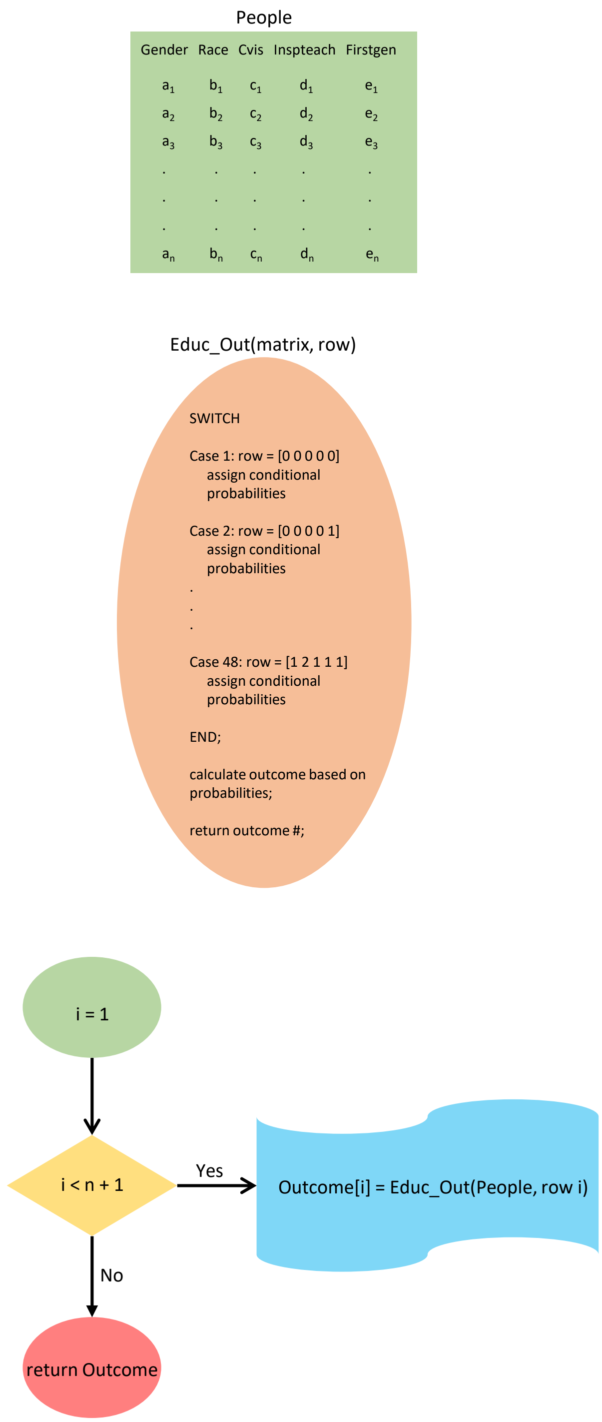

Figure 4.8: Flow diagram of the algorithm behind the Agent-Based Model.

4.5 White Female Model

After breaking down the data, the conditional probabilities for each level of schooling were computed within the white female population. Table 4.2 shows the responses for that data.

Educational Outcome	no_inspteach yes_firstgen	yes_inspteach yes_firstgen	no_inspteach no_firstgen	yes_inspteach no_firstgen
Elementary School	0	0	0	0
Middle School	0	0	0	0
High School	0	7	3	12
Associate	1	4	0	2
Bachelor	0	9	3	10
Master	1	1	2	18
Doctorate	0	0	0	5
Total	2	21	8	47

Table 4.2: Outcomes for white female survey participants.

One thing to note in Table 4.2 is that for someone to have achieved, say, their high school degree, they also would have had to complete elementary school and middle school. Table 4.3 reflects this difference with reverse cumulative columns. This case study is still relatively small, with only 2 and 8 subjects in the first and third columns. As a result, those conditional probabilities will not be very informative.

Educational Outcome	no_inspteach yes_firstgen	yes_inspteach yes_firstgen	no_inspteach no_firstgen	yes_inspteach no_firstgen
Elementary School	2	21	8	47
Middle School	2	21	8	47
High School	2	21	8	47
Associate	2	14	5	35
Bachelor	1	10	5	33
Master	1	1	2	23
Doctorate	0	0	0	5

Table 4.3: Cumulative outcomes for white female participants.

Table 4.3 will provide the information needed to calculate the conditional probabilities. For example, the conditional probability of getting a Bachelor's degree given the subject is first generation and had an inspirational teacher is calculated as:

$$\begin{aligned} P(bachelor \mid yes\_inspteach,\ yes\_firstgen) &= \frac{P(bachelor\ \cap\ yes\_inspteach,\ yes\_firstgen)}{P(yes\_inspteach,\ yes\_firstgen)} \\ &= \frac{\frac{10}{80}}{\frac{21}{80}} = 47.62\%. \end{aligned}$$

Note that this is higher than the national average for white females with Bachelor's degrees, which is 35.53%. The other conditional probabilities were calculated similarly. Once this was completed, these probabilities were entered into the ABM code. The exact code and conditional probabilities can be seen in Appendix C.

4.6 General Population Model

Due to the small sample size of the survey data collected, there was not enough information on each of the 48 combinations of people to calculate the needed conditional probabilities. To be able to see what a full model would look like given the proper data, the researchers decided to simulate their own data and form a model from that data. The simulated data is designed to mimic the population proportions seen in the Census data. Of the 500,000 simulated people, about 50% are female, and about 79% are white and 11% are black. The educational outcomes are designed to mimic the national distribution as well, though these are not based on any of the other attributes of the simulated people. This means that every simulated person has the same likelihood of educational attainment outcomes. Because of this, any advantages seen in the results are randomly assigned by the program and do not give any real-life insight into the interventions.

For the interventions, the population proportions from the survey data are used. About 24% of simulated people are assigned yes_cvis, 88% assigned to yes_inspteach, and 32% assigned to yes_firstgen. These assignments are applied randomly, so the exact proportions of the simulated data are a bit off from the population proportions.

Once this simulated data is made, the program goes through and computes the conditional probabilities for each level of educational attainment for each person combination, creating a total of 336 conditional probabilities. This project will look at one such set of simulated data. Shown below are the conditional probabilities for white females. The other 40 person combinations were also calculated, but they are not shown.

Educational Attainment	Elem. School	Middle School	High School	Associate Degree	Bachelor's Degree	Master's Degree	Doctorate Degree
Person							
[1 0 0 0 0]	0.9938	0.9814	0.8995	0.4552	0.3499	0.1348	0.0239
[1 0 0 0 1]	0.9938	0.9807	0.9031	0.4540	0.3529	0.1405	0.0262
[1 0 0 1 0]	0.9939	0.9802	0.9031	0.4562	0.3523	0.1326	0.0227
[1 0 0 1 1]	0.9933	0.9789	0.9016	0.4531	0.3515	0.1330	0.0238
[1 0 1 0 0]	0.9909	0.9775	0.9079	0.4224	0.3553	0.1324	0.0156
[1 0 1 0 1]	0.9909	0.9775	0.9079	0.4224	0.3553	0.1324	0.0156
[1 0 1 1 0]	0.9936	0.9787	0.8970	0.4618	0.3558	0.1362	0.0238
[1 0 1 1 1]	0.9935	0.9782	0.9036	0.4548	0.3533	0.1335	0.0242

Table 4.4: Example of the resulting conditional probabilities for white females. Computed with data taken directly from the survey.

As can be seen in Table 4.4, the probabilities are very similar across the different people. With real data, it would be expected that there would be more variation. Also note that [1 0 1 0 0] (female white yes_cvis no_inspteach not_firstgen) and [1 0 1 0 1] (female white yes_cvis no_inspteach firstgen) have the exact same probabilities . One issue that the model experienced was the rarity of that [x x 1 0 x] combination, which means that the subject did go on college visits and did not have an inspirational teacher. Even though the code generated 500,000 data points for this model, every run of the code ended up having some [x x 1 0 x] combination completely empty. In this case, the general cumulative probabilities found in Figure 3.1 were placed there instead, since the conditional probabilities could not be calculated for those empty subgroups. All of the conditional probabilities are then input into the function that will predict educational outcomes.

Results

5.1 White Female Model Results

The white female case study looks at the educational outcomes of 10,000 randomly generated people based on the collected survey data. In each case seen, a new matrix of people is generated for the specific question that is considered. For example, Figure 5.1 displays the educational outcomes of white females that received no interventions. This matrix of people is a random mix of first generation and not first generation. The output here serves as the control. The interpretation can be seen from the percentages of each educational outcome. About 45% of people achieved their high school degree but no higher. Only 23.48% of people achieved an Associate's Degree, but nearly 27% went on to achieve a Bachelor's Degree. Around 5% achieved a Master's Degree, but none of these people achieved a Doctorate, which comes directly from the survey data. The only subjects who received a Doctorate also said that they had an inspirational teacher/mentor, so the conditional probability of getting a doctorate while having no inspirational teacher is 0. This is one challenge of the small sample size.

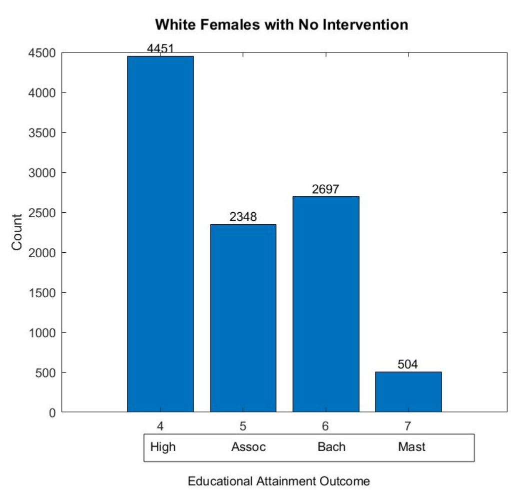

Figure 5.9: Results from running the model for white females with no inspirational teacher.

The first output to compare to the control case is the outcomes of a random mix of people: first generation with an inspirational teacher, not first generation with no inspirational teacher, etc. This can be seen in Figure 5.1. As previously noted, this output begins in high school which is to be expected since all of the females made it through high school in Table 4.3. This distribution is what may be seen in a general white female population, with a mix of people who are first generation and not, as well as a mix of people who received the inspirational teacher intervention and those who did not. This mixed matrix has higher percentages for every post-high school educational outcome. The Associate's Degree outcome increased by 3%, the Bachelor's Degree by 1%, and the Master's Degree by 3%. There is also the emergence of the Doctorate outcome, with about 0.71% of people achieving a this. Clearly, the introduction of the inspirational teacher intervention has made some positive impact on educational attainment.

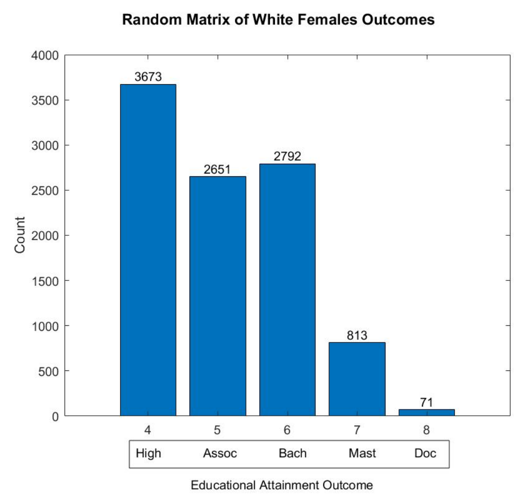

Figure 5.10: Results from running the model for a mix of the inspirational teacher intervention and parental educational attainment for white females.

The next step is to see the outcomes of a white female population (of mixed parental educational attainment) where every student has experienced the inspirational teacher intervention. This can be seen in Figure 5.1. The distribution between High School, Associate's, and Bachelor's is much more even than in 5.1. Giving this intervention to everyone appears to positively affect educational outcomes.

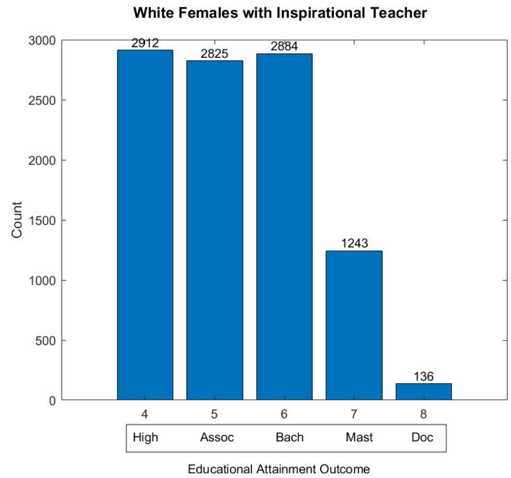

Figure 5.11: Results after running the model for white females all who had inspirational teachers.

Finally, the interaction of parental educational attainment and the inspirational teacher intervention can be studied. Figure 5.1 shows the case where all people are first generation and have received the inspirational teacher intervention, and Figure 5.1 shows the case where all people have at least one parent pursued higher education as well as receiving the inspirational teacher intervention. The first curious thing to note is that Figure 5.1 has nobody achieving a Doctorate, while about 2.5% achieve one in Figure 5.1. In this, it looks like parental educational attainment may have a large impact on terminal degrees. Furthermore, the Master's Degree receives a 22% increase from first generation to not first generation. It is apparent that the students who had at least one parent with a post-high school degree were more likely to achieve Bachelor's, Master's, and Doctorate Degrees.

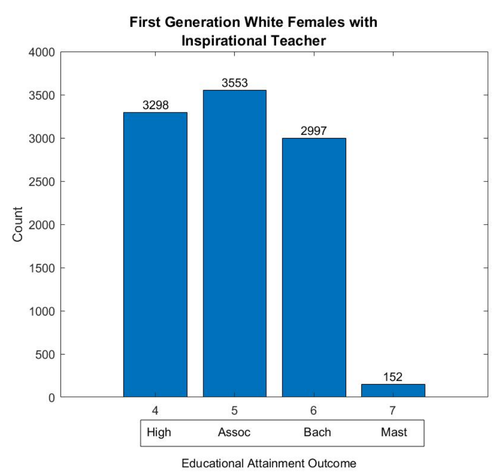

Figure 5.12: Results after running the model for first generation white females with inspirational teachers.

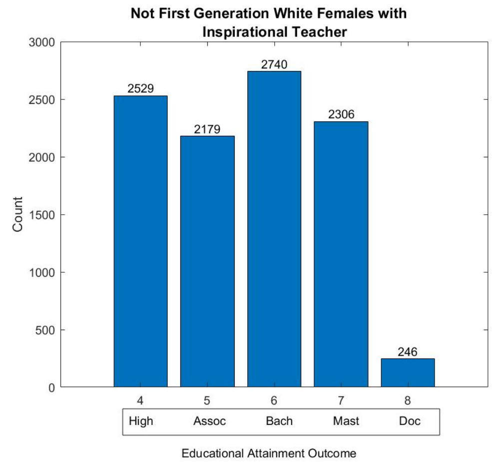

Figure 5.13: Results after running the model for not first generation white females with inspirational teachers.

From these plots, a few inferences can be made. First, having an inspirational teacher seems to improve educational outcomes, regardless of parental educational attainment. Furthermore, when combined with an inspirational teacher, parental educational attainment strongly impacts the educational attainment of the child. This would suggest that those who are first generation may need more guidance from the schools when it comes to pursuing higher education when compared to their peers who are not first generation. However, these outputs are based on a fairly small sample size, so more data would need to be collected to truly understand the impacts of these variables.

5.2 The General Model

The general model here is using the conditional probabilities from the generated data of the program discussed above. Gender, race, and parental education is randomly distributed along with the interventions. With a cursory run through of this model on a randomly generated matrix of 10,000 people, the educational attainment predictions are as seen in Figure 5.2. 51.09% of this population achieved a high school degree, but no higher, whereas only 0.05% of these people achieved a doctorate. This outcome shows what might be seen in any given school – some students have one or both interventions while some students do not have any.

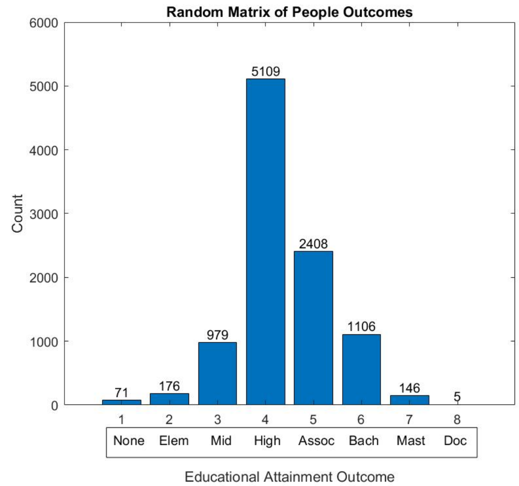

Figure 5.14: Results from running the model for all demographics with a mix of all interventions.

To better understand how the distribution of educational attainment is affected by interventions, Figure 5.2 shows the outcomes of a control group, that had neither the college visit intervention nor the inspirational teacher intervention. Once again, because the conditional probabilities are from generated data, the outputs here will not show any real relationships between the interventions and educational attainment outcomes. From the comparison between Figures 5.2 and 5.2, it appears that the control group actually has marginally better outcomes than the former. The only exceptions are the None and Doctorate educational outcomes.

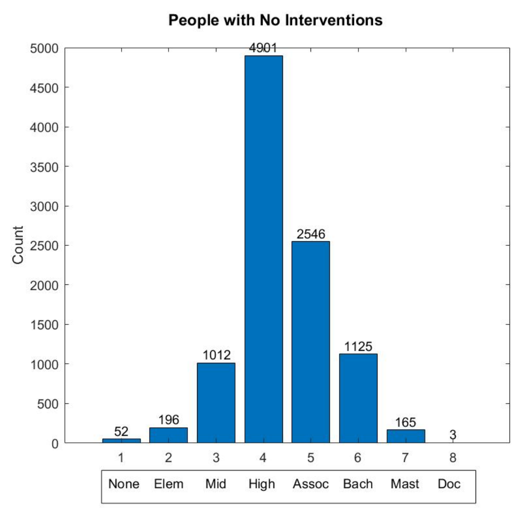

Figure 5.15: Results from running the model for all demographics with no interventions.

Figures 5.2 and 5.2 show the distributions if all people have the college visit intervention. Figure

5.2 shows the case where nobody has an inspirational teacher while Figure 5.2 shows the case where a random selection of people have an inspirational teacher. According to these graphs, it appears that the college intervention actually lowers the amount of students who go on to higher education when compared to the control group. In this case, use of college visits as an intervention would be discouraged for schools that wish to raise their educational attainment.

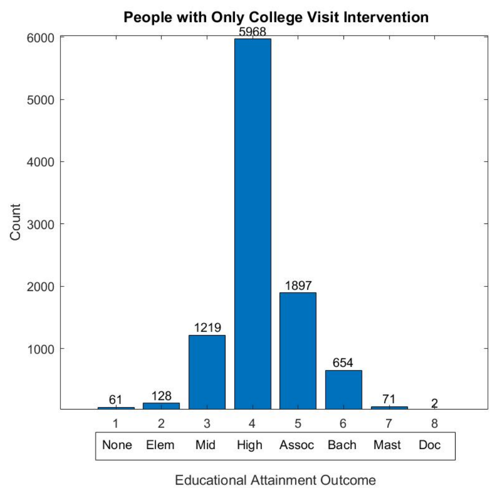

Figure 5.16: Results from running the model for all demographics with only the college visit intervention.

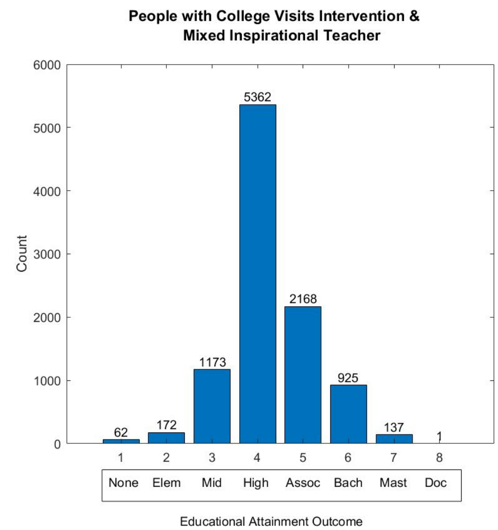

Figure 5.17: Results from running the model for all demographics with the college visit intervention and with some having an inspirational teacher and some without.

The next intervention to consider is that of inspirational teachers/mentors. As can be seen in Figures 5.2 and 5.2, there is the case where each student had only an inspirational teacher and the case where each student had an inspirational teacher, but a random selection also had the college visits intervention. Figure 5.2 has the best outcomes of any case seen from this model. About 26.7% of these people received an Associate's Degree, while 12%, 2%, and 0.07% received a Bachelor's, Master's, and Doctorate Degree respectively. All of these percentages are higher than the control group in Figure 5.2 and the random group in Figure 5.2. However, it appears that including the college visits intervention lowers those higher education percentages, meaning that for optimal educational outcomes, a school should only implement the inspirational teacher intervention.

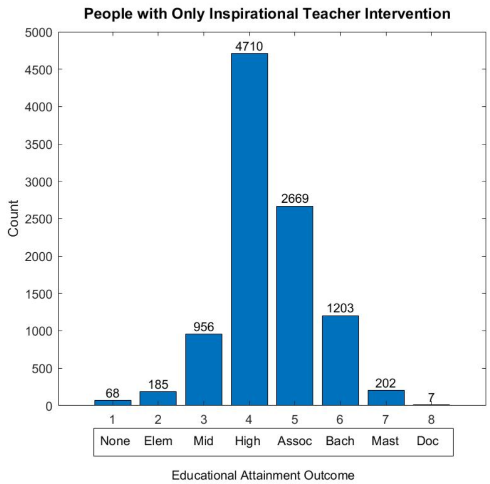

Figure 5.18: Results from running the model for all demographics with only the inspirational teacher intervention.

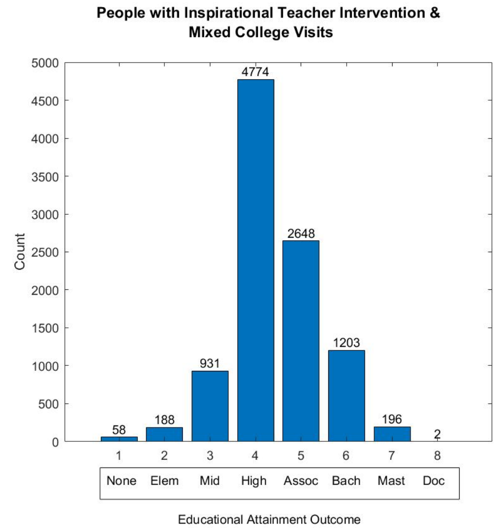

Figure 5.19: Results from running the model for all demographics with inspirational teacher intervention and some with the college visit intervention and some without.

Finally, the last case to consider is one in which every student has both an inspirational teacher and has participated in a school-organized college visit. This can be seen below in Figure 5.2. This actually appears to be less successful than 5.2, which is consistent with the apparent negative impact of the college visit intervention. The Doctorate count increases by 1, but that is too minimal to mean anything. Thus, it appears that the best action to take given the generated data is to only use the inspirational teacher intervention and disregard the college visits intervention.

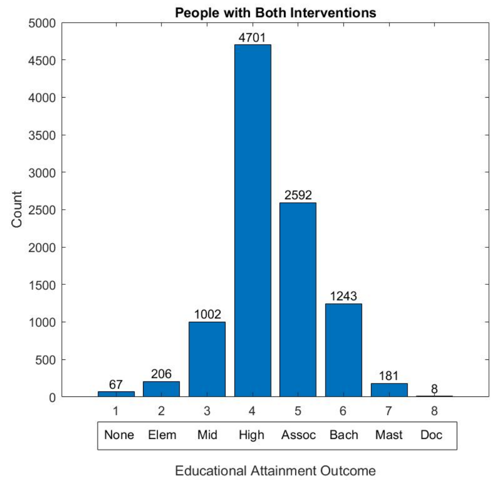

Figure 5.20: Results from running the model for all demographics with both the college visit intervention and the inspirational teacher intervention.

5.3 Conclusions

From the true survey data seen with the white female case study, the inspirational teacher intervention is found to be a strong tool when it comes to improving educational attainment. The outputs seen with this model suggest that schools would be wise to inform teachers of their strong impact and encourage those teachers to reach out to students who are struggling. Given enough survey data, these outcomes could be expanded to include other interventions, as well as the interaction between the many recorded factors. Meanwhile, the general model that is formed from generated data shows a glimpse of that scope, with the study of the inspirational teacher and the college visit interventions. While the inspirational teacher intervention improved the educational outcomes, the college visit intervention ended up decreasing the educational outcomes. Of all possible combinations of interventions, the best was getting every student to have an inspirational teacher. This relationship would probably be different given real data, but it illustrates the interesting interplay between interventions.

Future Works

Agent-based models are some of the most build-able models available. That means that there is an extensive amount of information that can be added and changed to build an even more illuminating model. As a result, this research has only scratched the surface of potential that is available to this project.

The first step to continuing this work is to collect more data. This can be done by refiling with the Youngstown State Institutional Review Board so that the survey could offer a prize draw. Since motivating the responses was the most difficult part, offering a reward might help garner more responses. Another option could be to find business sponsors, such as restaurants or stores, who would offer discounts to customers who showed proof of taking the survey. This would be a good possibility in small towns, and it would be able to reach a wider audience than that of the original survey.

Additionally, the survey questions can be expanded to add more possible directions to take the research. One factor that was not considered during the first round of data collection is the impact of peer support. Asking questions such as:

- What level of education did your closest high school friend achieve?
- What level of impact did your friends' educational decisions have on your own?
- Did you have a friend who encouraged you to pursue higher education?
- Did you ever encourage one of your friends to pursue higher education?

would introduce an agent-to-agent interaction into the model. Using that information, if the agent in question was in close enough proximity to an "encouraging" agent, then its conditional probabilities could be increased. With this, the model could find the optimal number of "encouraging" agents a certain school would need to have good improvement in educational attainment. The friendship factor could turn out to be even more influential than the original factors recorded in the survey, as teenagers tend to listen to their peers more than authority figures.

Another important step to take is to perform statistical analysis on the collected data and perhaps fit a multiple linear regression model or an ordinal regression model to it. A sample size of 128 is enough data to build a decent regression model. This would provide more insight into the data and allow for the interaction between interventions to come to light.

Furthermore, the ultimate goal is to have this model applied to a specific high school to determine which interventions would be most effective. The model could be run on a set of people that demographically aligns with the student population. After running it with no interventions and recording that output, researchers could then run it with different combinations of interventions to see which maximize the educational attainment outcomes. These interventions could then be applied in to that actual high school. If a high school were to partner with this research, then they might be able to offer data on how those interventions make a difference in their student body.

Bibliography

[1] C. R. Belfield and H. M. Levin, eds., *The Price We Pay: Economic and Social Consequences of Inadequate Education*, Brookings Institiution Press, 2007.

[2] E. F. Dubow, P. Boxer, and L. R. Huesmann, *Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations*, Merrill-Palmer Quarterly (Wayne State University), 55 (2009), p. 224–249.

[3] C. M. Macal and M. J. North, *Tutorial on agent-based modelling and simulation*, Journal of Simulation, 4 (2010), pp. 151–162.

[4] Magnet Schools of America, *What are magnet schools*. `https://magnet.edu/about/what-are-magnet-schools#1499667889100-039b81ce-813c`, 2021.

[5] R. L. Ott and M. Longnecker, *An Introduction to Statistical Methods and Data Analysis*, Cengage Learning, 7 ed., 2018.

[6] S.-R. Ou, *Pathways of long-term effects of an early intervention program on educational attainment: Findings from the chicago longitudinal study*, Journal of Applied Developmental Psychology, 26 (2005), pp. 578–611.

[7] C. L. Ryan and K. Bauman, *Educational attainment in the united states: 2015*, 2016.

[8] A. Saleh and K. Bista, *Examining factors impacting online survey response rates in educational research: Perceptions of graduate students*, Journal of MultiDisciplinary Evaluation, 13 (2017), p. 63–74.

[9] E. Schreiber, *barvalues*(*h, precision, textparams*), 2022.

[10] United States Census Bureau, *Cps historical time series visualizations*. `https://www.census.gov/library/visualizations/time-series/demo/cps-historical-time-series.html`, 2021.

[11] ———, *Educational attainment in the united states: 2018*, October 2021.

[12] ———, *Quickfacts: Ohio*. https://www.census.gov/quickfacts/fact/table/OH/PST045219, 2021.

Appendices

Appendix A

Survey Questions

1. Did you do K-12 education entirely in Ohio?
2. Please list the school district name where you completed your education below:

   The following questions will ask you to rate the impact certain experiences had on your decision to pursue higher education. They are listed and explained as follows:

   - 1 – minimal impact: This experience had nothing to do with my decision. I would have made the same decision regardless of this experience.
   - 2 – medium impact: This experience definitely strengthened my decision. Without this experience, my decision may have been different.
   - 3 – high impact: This experience was one of the sole reasons for my decision. My decision would have been different without this experience.
3. While in K-12 schooling, did you take any college preparatory courses?
   (a) Please list all college preparatory courses you have taken.
   (b) Please rate the impact of these courses on your decision regarding whether or not to pursue higher education.
4. While in K-12 schooling, did you go on any college visits organized by your school district?
   (a) Please rate the impact of these visits on your decision regarding whether or not to pursue higher education.

5. While in K-12 schooling, did you have an inspiring teacher/mentor somewhere in your schooling?
   (a) Please rate the impact of this teacher/mentor on your decision regarding whether or not to pursue higher education.
6. While in K-12 schooling, were you encouraged by any faculty to pursue higher education?
   (a) Please rate the impact of this encouragement on your decision regarding whether or not to pursue higher education.
7. While in K-12 schooling, did you participate in any school sports?
   (a) Please list the sports you participated in.
   (b) Please rate the impact of this/these sports on your decision regarding whether or not to pursue higher education.
8. While in K-12 schooling, did you participate in any after school academic clubs?
   (a) Please list the activity(ies) you participated in.
   (b) Please rate the impact of this/these activities on your decision regarding whether or not to pursue higher education.
9. While in K-12 schooling, did you participate in any other extra-curricular activities?
   (a) Please list the activity(ies) you participated in.
   (b) Please rate the impact of these activities on your decision regarding whether or not to pursue higher education.
10. What is your current occupation?
11. What was the most positive impact on your academic career?
12. What was the most negative impact on your academic career?
13. What is your age (in years)?
14. What is your gender?
15. What is your race?
16. What is the highest level of education YOU have completed?

17. What is the highest level of education your MOTHER has completed?

18. What is the highest level of education your FATHER has completed?

19. What is the highest level of education your CHILD has completed?

Appendix B

IRB Waiver Form

Youngstown State University
Institutional Review Board
Office of Research
Melnick Hall
Phone: 330-941-2377

4/2/2021 — Date

090-21 — Protocol Number

FULL/EXPEDITED REVIEW PROTOCOL APPLICATION

A. INVESTIGATOR INFORMATION			
Please list all study personnel involved in the conduct of this study. All study personnel must complete required training in human subject research and provide to the IRB office documentation verifying completion of the requirement. The IRB will not review a study without such forms on file for all research personnel. Only YSU faculty, staff, students, or registered volunteers are considered YSU affiliated and thus covered by the YSU IRB review. All non-affiliated study personnel must have their participation reviewed by the appropriate IRB. (Attach a separate sheet if more space is needed.)			
STUDY TITLE	A Mathematical Model for Educational Attainment		
PRINCIPAL INVESTIGATOR OR FACULTY ADVISOR	Dr. Alicia Prieto-Langarica	Phone Extension (330) 941-1549	Email Address aprietolangarica@ysu.edu
DEPARTMENT	Mathematics and Statistics		
CO-INVESTIGATOR OR STUDENT INVESTIGATOR	Anna Truman	Phone Extension (330) 936-8057	Email Address actruman@student.ysu.edu
CO-INVESTIGATOR OR STUDENT INVESTIGATOR		Phone Extension	Email Address
CO-INVESTIGATOR OR STUDENT INVESTIGATOR		Phone Extension	Email Address

B. SPONSOR/FUNDING INFORMATION			
Will this project be supported by an external funding agency?	☐ Yes	☑ No	
If yes, please identify the source and contact information			
Agency:	Contact Person:	Phone:	Email:

C. LOCATION OF RESEARCH			
Where will the study take place?	YSU	Other Facility	
If not at YSU, attach a letter of cooperation on the letterhead of the facility and provide contact information. If there are multiple facilities, attach an additional page with the information for each.			
Facility Name:	Contact Person:	Phone:	Email:

D. CONFLICT OF INTEREST	
Is there any real or apparent conflict of interest on the part of any study personnel (e.g., investigator/participant relationship, stock or stock options, interest in technology, consultant to sponsor)?	☐ Yes ☑ No
If yes, please explain.	

E. METHODS AND PROCEDURES
This section must be written in layman's terms so that it can be understood by all members of the IRB. Include sufficient detail so that reviewers will be able to fully understand your project.
1. Describe the background of the research and the significance of the study. In today's society of high-stress and high-performance academics, many students are left behind in the race to climb the educational ladder. Whether from lack of support, lack of exposure, or lack of opportunity, there are certain demographics that have lower educational attainment than their peers. This research aims to discover students who lack this support and determine what interventions would be most effective to give them equal exposure to higher education opportunities. With this study, the researchers hope to provide an objective and mathematical overview/solution of this societal inequality.
2. What is the objective of the study? The objective of this study is to determine what factors are important indicators of educational attainment levels and to build a mathematical model that estimates the average educational attainment for a certain subject. After discovering these factors, the study will aim to try different interventions for each subject and see which interventions appear to be most effective in increasing educational attainment.
3. Describe the study design and all procedures (sequentially) to which human subjects will be exposed. Human subjects will only be exposed to a questionnaire to serve as data collection. This questionnaire will ask basic demographic questions, questions about educational attainment, and questions about past educational experiences. This data will then be recorded and uploaded to be used as a base for the model.
4. If deception is to be used in this study, describe and justify the deception and explain the debriefing procedures. No deception is to be used.
5. Will subjects be presented with materials that they might regard to be offensive, threatening, or degrading? No.
6. Reference pertinent scientific literature. The U.S. Census Bureau released an article in March 2020 from the 2019 Educational Attainment data set. In this article, they discuss the different percentages of educational attainment. "In 2019, 40.1% of non-Hispanic whites age 25 and older had a bachelor's degree or higher, up from 33.2% in 2010. During the same period, the percentage of blacks age 25 and older with a bachelor's degree or higher rose from 19.8% to 26.1%; Asians from 52.4% to 58.1%; and Hispanics from 13.9% to 18.8%."[1] From this data, we can see that there are certainly disparities between different races when it comes to educational attainment. More information from the U.S. Census Bureau is available that shows additional inequalities for other factors. This, and many similarly reported statistics, serves as inspiration for this research. [1] Source from https://www.census.gov/newsroom/press-releases/2020/educational-attainment.html#:~:text=In%202019%2C%20high%20school%20was,from%2029.9%25%20to%2036.0%25

F. SURVEYS AND QUESTIONNAIRES, IF APPLICABLE		
☑ Survey/Questionnaire (go to A)	☑ Record or Database (go to B)	☐Other, Briefly Explain

A.	**Surveys and Questionnaires. You must attach a copy of each survey or questionnaire**.
1.	What type of survey or questionnaire will be used? Online questionnaire.
2.	Describe the setting and mode of administration for the instrument (e.g. by phone, on-on-one, group) and the provisions for maintaining privacy and confidentiality (e.g. anonymity). Include duration, intervals of administration, and overall length of participation. This will be an online questionnaire distributed over social media. It will be about a 10-minute survey that can be completed all at once. Overall length of participation should be around 10 minutes.
B.	**Records or Data Review. This includes existing material such as archival records, databases, etc.**
1.	What kinds of records will you review? What is the source of the records? The records will be state/government recorded data. Data will be obtained from the United States Census Bureau, the Ohio Department of Education, and the Ohio Department of Higher Education.
2.	Will you have contact or interaction with the subjects from whom the data are collected? There will be no interaction with the subjects of the data.
3.	Will you be recording identifiers (information that could potentially identify human subjects)? We will only be recording demographics as supplied by the data – no unique identifiers would be used.
4.	Define the time frame of the records that you plan to review. (Example: from 2/1/2007-2/1/2008) Data ranges from 2000 – 2019.

G. RISK/BENEFIT ASSESSMENT		
Describe in detail any potential risks/adverse events associated with each research procedure.		
1. Determine the level of risk to subjects associated with this project.		
☑ None	☐ Minimal	☐ Low, moderate, high
DHHS and FDA Regulations define minimal risk as "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests."		
2. Describe in detail any potential risks/adverse events that could be associated with the research procedures Students will fill out a basic demographic and educational attainment questionnaire. This will not lead to any risks for the students, as it will be anonymous and objective.		
3. Describe the potential direct benefits subjects may receive as a result of their participation. Subjects will not receive any direct benefits from this study.		
4. Describe any potential benefits that may accrue to the subject or society. *Societal benefits generally refer to the advancement of scientific knowledge, and/or possible benefit(s) to future subjects.*		

G. RISK/BENEFIT ASSESSMENT
With this study, we will be able to theorize college preparation programs and interventions aimed at certain demographics to raise educational attainment. In this way, future students have the potential benefit of increased intervention to open pathways to college and other educational opportunities.
5. Explain how the benefits of this research outweigh the potential risks and how these risks are justified. Since this research would not require any risk, the potential benefits of this study would outweigh the risks.

H. HUMAN RESEARCH SUBJECTS		
Describe the target population in specific terms.		
1. Provide detail about the numbers of subjects, age, gender, or any other information that establishes the parameters of the population in your study. Because the population that we're studying is Ohio, our sample will mimic the different factor distributions of the Ohio population. We hope to get a response of at least 1,000. The acceptable ages will be 18 and over.		
2. Outline the criteria for selection and exclusion of subjects. Subjects will be selected randomly from adults that attended school in Ohio.		
3. Approximately how much time will be required of each subject? About 10 minutes to fill out the survey.		
4. Describe the different conditions or manipulations to be conducted in the study. None.		
5. Will any of the following vulnerable populations be targeted for subject recruitment?		
☐ Minors	☐ Mentally incapacitated	☐
☐ Prisoners	☐ Elderly	☐
☐ Pregnant women/fetuses	☐ Non-English speaking	☐
6. What safeguards are in place to protect vulnerable populations if involved within the research? N/A		
7. Describe the measures or observations that will be taken in the study. Level of education, demographics, and other details related to schooling experience.		
8. What steps will be taken to ensure that subject's participation is voluntary? The questionnaire will be completely voluntary and no single person will ever be asked to complete it.		
9. Will the subjects receive compensation for their participation, monetary or otherwise? If so, please specify. No.		
10. What financial obligations will subjects incur as a result of participating in this research? Identify expenses such as travel costs, parking fees, missed work, etc. Please be as specific as possible. No financial obligations will be incurred.		

I. RECRUITMENT PROCEDURES
1. What method(s) will be used to identify and recruit prospective Subjects? Specify the source of potential subjects. Survey will be distributed through public social media groups based in Ohio.
2. Will you access existing stored data, records, etc. for your recruitment purposes? If yes, specify the source. Yes. The records will be state/government recorded data. Data will be obtained from the United States Census Bureau, the Ohio Department of Education, and the Ohio Department of Higher Education.

J. INFORMED CONSENT AND ASSENT
Ethical and regulatory guidelines ensure that potential subjects must be fully informed about the research in a manner comprehensible to them and then be allowed to choose whether to participate in the research. Attach an Informed Consent Form of your own design, according to the YSU Guidelines for fully Informed Consent for each subject population, or a Waiver of Informed Consent Request Form. The IRB has provided a template containing the Elements of Informed Consent/Assent (per 45 CFR 116) on the YSU IRB website: http://cms.ysu.edu/administrative-offices/research/human-subjectsinstitutional-review-board. Using the template is strongly suggested in order to eliminate errors and revisions.
Select only one of the three boxes below:
☐ **I am attaching a copy of all Consent and Assent forms that will be used in this study and will answer the questions below.** *A letter of consent is generally required form all adult research participants unless specifically waived by the IRB. A letter of assent is required of all minor research participants (age 9-17) unless specifically waived by the IRB.*
☐ **I am requesting that the IRB allow me to use an oral consent process for my research and will answer the questions below.** The elements of informed consent may be presented orally to the subject or the subject's legally authorized representative. When this method is used there shall be: 1. A witness to the oral presentation. 2. The IRB shall approve a written summary of what is to be said to the subject or the representative. 3. A short form stating that the Elements of Informed Consent required by Section 45.116 have been presented orally to the subject or the subject's legally authorized representative and must be signed by the subject or the representative. 4. The witness shall sign both the short form and a copy of the summary. 5. A copy of the summary shall be given to the subject or the representative, in addition to a copy of the short form. **State specifically why you are asking the IRB to allow you to waive written consent and use an oral consent process.**
☑ I am requesting that the IRB waive my requirement for a signed letter of consent/assent. A signed consent form may be waived if the IRB finds either: 1. That the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality. Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern; or 2. The research presents no more that minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context. **State specifically why you are asking the IRB to waive the requirement for a signed letter of consent/assent.** This research requires no risk for the subject. No identifiers will be collected, as we are purely interested in the demographics data. The questionnaire will simply link the different demographics to a single subject (like race, gender, school district, etc.) to allow for compounded probabilities to be computed. Each research subject will be anonymous when mixed in with the collected data. The point of this research is not to report the results of the questionnaire, but to use the questionnaire to build a realistic model of different students. For example, suppose one subject is a white low-

J. INFORMED CONSENT AND ASSENT
class male in the Akron City School district. Whatever his highest level of educational attainment is will be recorded and used in concordance with all other white low-class males in the same district. The average probability will be used in the model to build a simulated sample of thousands of students, and the results of that model will be reported. This individual's information will never be reported.
Answer All of the Questions Below
1. How and where will informed consent/assent be obtained? (e.g., in the school, Investigator's office, etc.) N/A
2. Will there be an opportunity for potential subjects to take the consent form home to consider the options and to discuss participation with family members. If not, explain why. N/A
3. If subjects are minors or mentally disabled, describe how and from whom permission will be granted? N/A
4. How and by whom will it be determined that the subjects or their legal representative understand the research project and their rights as participants? N/A
5. Where will the record(s) of consent/assent be stored? N/A

K. CONFIDENTIALITY OF INFORMATION COLLECTED
1. What steps will be taken to ensure the anonymity or confidentiality of the subjects' identities or the data they provide. The collected data will include no identifiers or names. The data will only be seen by the researchers listed on the project or by other academics brought in to consult by Dr. Prieto. The reports will never include a questionnaire response – the data will only be used to build an agent-based model and to test the efficacy of the model.
2. Explain how the data will be stored, for what period of time, and how and when it will be disposed of. The collected data will be stored on the researcher's private computers. There are no plans to dispose of the data since the model will likely require many improvements that rely on information found in the data.

L. ASSURANCES – Principal Investigator or Faculty Advisor AND Student Investigator

PRINCIPAL INVESTIGATOR'S ASSURANCE STATEMENT

I certify that the information provided in this claim of exemption is complete and correct.

I understand that as Principal Investigator, I have the ultimate responsibility for the protection of the rights and welfare of human subjects and the ethical conduct of this research protocol. I agree to comply with all IRB and Institutional policies and procedures, as well as with all applicable federal, state, and local laws regarding the protection of human subjects in research, including, but not limited to, the following:

- The research will not be initiated until written approval is secured from the IRB.
- The project will be performed by qualified personnel according to the research protocol

Statement of Consent included at beginning of survey:

I am a graduate student at Youngstown State University and am conducting a research project as part of my graduate fellowship program. This research aims to discover educational inequalities in the Ohio schooling system. By collecting demographic and schooling data, we will build what is called an Agent-Based Model (ABM) to model the educational path of a student of certain parameters. The goal is to determine what types of students may lack educational support and to find what interventions would be most effective to give them equal exposure to higher education opportunities. With this study, we hope to provide an objective and mathematical overview/solution of these societal inequalities. You must be over the age of 18 to participate, and your participation is completely voluntary. Participation is expected to take between 5 and 15 minutes. There will be no consequences if you choose not to participate and you may withdraw at any time. There is no known risk to you. Your participation in this study will lead to a better understanding of Ohio's educational system and a new view on interventions, so that subsequent generations may be offered more equal opportunities. All information you provide will be kept completely anonymous and you will not be able to be identified when reported. Completion of this survey indicates your consent to participate. If you have any questions about the research project please contact Dr. Alicia Prieto-Langarica at aprietolangarica@ysu.edu. If you have any questions regarding your right as a research participant, please contact the Office of Research Services at 330-941-2377 or YSU IRB@ysu.edu.

Approval email from Karen H. Larwin, Friday, April 2, 2021

Dear Investigators,

Your protocol entitled A Mathematical Model for Educational Attainment has been reviewed and is deemed to meet the criteria of an exempt protocol. You will be surveying adult over social media about their K-12 experiences and educational goals. Participants will be asked to complete a basic demographic information and several items covering questions about their education. The survey will be disseminated electronically. No identifying information is being collected.

The research project meets the expectations of 45 CFR 46.104(b)(2) and is therefore approved. You may begin the investigation immediately. Please note that it is the responsibility of the principal investigator to report immediately to the YSU IRB any deviations from the protocol and/or any adverse events that occur. Please reference your protocol number 090-21 in all correspondence about the research associated with this protocol.

Appendix C

MATLAB Code

C.1 White Female Educational Outcome Model

```
function [outcome] = White_Female_Model(M,rowNumber)

Person=M(rowNumber,:);                    %Creates a 1x5 matrix of one person from matrix M

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Value   Gender    Race      CVisits    InspTeach    ParentEd
%----------------------------------------------------------------------
% 0       Male      White     No         No           First gen
% 1       Female    Black     Yes        Yes          Not first gen
% 2                 Other
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 25: [female White no_cvisits no_inspteach first_gen]
if Person == [1 0 0 0 0]
    elem=1;
    mid=1;
    high=1;
    assoc=.5;
    bach=.5;
    mast=0;
    phd=0;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 26: [female White no_cvisits no_inspteach not_first_gen]
if Person == [1 0 0 0 1]
    elem=1;
    mid=1;
    high=1;
    assoc=.625;
    bach=.625;
    mast=.25;
    phd=0;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 27: [female White no_cvisits inspteach first_gen]
if Person == [1 0 0 1 0]
    elem=1;
    mid=1;
    high=1;
    assoc=.667;
    bach=.4762;
    mast=.0476;
    phd=0;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 28: [female White no_cvisits inspteach not_first_gen]
if Person == [1 0 0 1 1]
    elem=1;
    mid=1;
    high=1;
    assoc=.7447;
    bach=.7021;
    mast=.4894;
    phd=.1064;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

a=zeros(1,8);                   %Creates array of zeros with specified size
[rows, columns]=size(a);

a(1)=1;                         %Represents the individual who will
                                %move through the levels of educ
                                %based on their assigned values
if rand < elem
    a(1)=a(1)-1;
    a(2)=a(2)+1;

    if rand < mid
        a(2)=a(2)-1;
        a(3)=a(3)+1;

        if rand < high
```

```
            a(3)=a(3)-1;
            a(4)=a(4)+1;

            if rand < assoc
                a(4)=a(4)-1;
                a(5)=a(5)+1;

                if rand < bach
                    a(5)=a(5)-1;
                    a(6)=a(6)+1;

                    if rand < mast
                        a(6)=a(6)-1;
                        a(7)=a(7)+1;

                        if rand < phd
                            a(7)=a(7)-1;
                            a(8)=a(8)+1;
                        end
                    end
                end
            end
        end
    end
end

if a(1)==1
    outcome = 1; % did not finish any schooling
elseif a(2) ==1
    outcome = 2; % only completed elementary
elseif a(3)==1
    outcome = 3; % only completed middle school
elseif a(4)==1
    outcome = 4; % only completed high school
elseif a(5) ==1
    outcome = 5; % only completed associate's
elseif a(6)==1
    outcome = 6; % only completed bachelor's
elseif a(7)==1
    outcome = 7; % only completed master's
elseif a(8)==1
    outcome = 8; % completed phd
else
    outcome = 0;

end
```

C.2 Running Results of White Female

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Value   Gender    Race      CVisits   InspTeach   ParentEd
%-------------------------------------------------------------------------------
% 0       Male      White     No        No          First gen
% 1       Female    Black     Yes       Yes         Not First Gen
% 2                 Other
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

numppl = 100000;

C1 = randi([1 1],[numppl 1]); % all females
C2 = randi([0 0],[numppl 1]); % all white
C3 = randi([0 0],[numppl 1]); % all nocvis
C4 = randi([0 1],[numppl 1]);
C5 = randi([0 1],[numppl 1]);

People = horzcat(C1,C2,C3,C4,C5);
Insp_Firstgen = [1 0 0 1 1];
Insp_Notfirstgen = [1 0 0 1 1];
Noinsp_Firstgen = [1 0 0 0 1];
Noinsp_Notfirstgen = [1 0 0 0 1];

nnz(ismember(People,Insp_Firstgen,'rows'))
nnz(ismember(People,Insp_Notfirstgen,'rows'))
nnz(ismember(People,Noinsp_Firstgen,'rows'))
nnz(ismember(People,Noinsp_Notfirstgen,'rows'))

[R, C]=size(People);

outcomeMatrix = zeros(R,1);

for i=1:R
    outcomeMatrix(i) = White_Female_Model(People,i);
end

B = unique(outcomeMatrix);
out = [B,histc(outcomeMatrix,B)];

bar(out(:,2))
set(gca,'xticklabel',{out(:,1)})
barvalues;
```

C.3 General Educational Outcome Model

```
%Singular Model 2 will contain every combination of variables
%a person can be.

function [outcome] = Singular_Model_3(M,rowNumber,A)

Person=M(rowNumber,:);                 %Creates a 1x5 matrix of one
        person from matrix M

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Value   Gender    Race      CVisits   InspTeach   ParentEd
%-------------------------------------------------------------------------------
% 0       Male      White     No        No          First gen
```

```
% 1        Female    Black        Yes        Yes        Not first gen
% 2                  Other
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Calculating outputs for people in input matrix.
elem=0;
mid=0;
high=0;
assoc=0;
bach=0;
mast=0;
phd=0;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 1: [male White no_cvisits no_inspteach first_gen]
if Person == [0 0 0 0 0]
    matrix = A{1};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 2: [male White no_cvisits no_inspteach not_first_gen]
if Person == [0 0 0 0 1]
    matrix = A{2};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 3: [male White no_cvisits inspteach first_gen]
if Person == [0 0 0 1 0]
    matrix = A{3};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 4: [male White no_cvisits inspteach not_first_gen]
if Person == [0 0 0 1 1]
    matrix = A{4};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 5: [male White cvisits no_inspteach first_gen]
if Person == [0 0 1 0 0]
    matrix = A{5};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 6: [male White cvisits no_inspteach not_first_gen]
if Person == [0 0 1 0 1]
    matrix = A{6};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 7: [male White cvisits inspteach first_gen]
if Person == [0 0 1 1 0]
    matrix = A{7};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 8: [male White cvisits inspteach not_first_gen]
if Person == [0 0 1 1 1]
    matrix = A{8};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 9: [male Black no_cvisits no_inspteach first_gen]
if Person == [0 1 0 0 0]
    matrix = A{9};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 10: [male Black no_cvisits no_inspteach not_first_gen]
if Person == [0 1 0 0 1]
    matrix = A{10};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
```

```
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 11: [male Black no_cvisits inspteach first_gen]
if Person == [0 1 0 1 0]
        matrix = A{11};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 12: [male Black no_cvisits inspteach not_first_gen]
if Person == [0 1 0 1 1]
        matrix = A{12};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 13: [male Black cvisits no_inspteach first_gen]
if Person == [0 1 1 0 0]
        matrix = A{13};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 14: [male Black cvisits no_inspteach not_first_gen]
if Person == [0 1 1 0 1]
        matrix = A{14};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 15: [male Black cvisits inspteach first_gen]
if Person == [0 1 1 1 0]
        matrix = A{15};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 16: [male Black cvisits inspteach not_first_gen]
if Person == [0 1 1 1 1]
        matrix = A{16};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 17: [male Other no_cvisits no_inspteach first_gen]
if Person == [0 2 0 0 0]
        matrix = A{17};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 18: [male Other no_cvisits no_inspteach not_first_gen]
if Person == [0 2 0 0 1]
        matrix = A{18};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 19: [male Other no_cvisits inspteach first_gen]
if Person == [0 2 0 1 0]
        matrix = A{19};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 20: [male Other no_cvisits inspteach not_first_gen]
if Person == [0 2 0 1 1]
        matrix = A{20};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 21: [male Other cvisits no_inspteach first_gen]
if Person == [0 2 1 0 0]
        matrix = A{21};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
```

```
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 22: [male Other cvisits no_inspteach not_first_gen]
if Person == [0 2 1 0 1]
        matrix = A{22};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 23: [male Other cvisits inspteach first_gen]
if Person == [0 2 1 1 0]
        matrix = A{23};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 24: [male Other cvisits inspteach not_first_gen]
if Person == [0 2 1 1 1]
        matrix = A{24};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 25: [female White no_cvisits no_inspteach first_gen]
if Person == [1 0 0 0 0]
        matrix = A{25};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 26: [female White no_cvisits no_inspteach not_first_gen]
if Person == [1 0 0 0 1]
        matrix = A{26};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 27: [female White no_cvisits inspteach first_gen]
if Person == [1 0 0 1 0]
        matrix = A{27};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 28: [female White no_cvisits inspteach not_first_gen]
if Person == [1 0 0 1 1]
        matrix = A{28};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 29: [female White cvisits no_inspteach first_gen]
if Person == [1 0 1 0 0]
        matrix = A{29};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 30: [female White cvisits no_inspteach not_first_gen]
if Person == [1 0 1 0 1]
        matrix = A{30};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 31: [female White cvisits inspteach first_gen]
if Person == [1 0 1 1 0]
        matrix = A{31};
        elem=matrix(2);
        mid=matrix(3);
        high=matrix(4);
        assoc=matrix(5);
        bach=matrix(6);
        mast=matrix(7);
        phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 32: [female White cvisits inspteach not_first_gen]
if Person == [1 0 1 1 1]
        matrix = A{32};
        elem=matrix(2);
        mid=matrix(3);
```

```
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 33: [female Black no_cvisits no_inspteach first_gen]
if Person == [1 1 0 0 0]
    matrix = A{33};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 34: [female Black no_cvisits no_inspteach not_first_gen]
if Person == [1 1 0 0 1]
    matrix = A{34};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 35: [female Black no_cvisits inspteach first_gen]
if Person == [1 1 0 1 0]
    matrix = A{35};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 36: [female Black no_cvisits inspteach not_first_gen]
if Person == [1 1 0 1 1]
    matrix = A{36};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 37: [female Black cvisits no_inspteach first_gen]
if Person == [1 1 1 0 0]
    matrix = A{37};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 38: [female Black cvisits no_inspteach not_first_gen]
if Person == [1 1 1 0 1]
    matrix = A{38};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 39: [female Black cvisits inspteach first_gen]
if Person == [1 1 1 1 0]
    matrix = A{39};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 40: [female Black cvisits inspteach not_first_gen]
if Person == [1 1 1 1 1]
    matrix = A{40};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 41: [female Other no_cvisits no_inspteach first_gen]
if Person == [1 2 0 0 0]
    matrix = A{41};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 42: [female Other no_cvisits no_inspteach not_first_gen]
if Person == [1 2 0 0 1]
    matrix = A{42};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 43: [female Other no_cvisits inspteach first_gen]
if Person == [1 2 0 1 0]
    matrix = A{43};
```

```
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 44: [female Other no_cvisits inspteach not_first_gen]
if Person == [1 2 0 1 1]
    matrix = A{44};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 45: [female Other cvisits no_inspteach first_gen]
if Person == [1 2 1 0 0]
    matrix = A{45};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 46: [female Other cvisits no_inspteach not_first_gen]
if Person == [1 2 1 0 1]
    matrix = A{46};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 47: [female Other cvisits inspteach first_gen]
if Person == [1 2 1 1 0]
    matrix = A{47};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Case 48: [female Other cvisits inspteach not_first_gen]
if Person == [1 2 1 1 1]
    matrix = A{48};
    elem=matrix(2);
    mid=matrix(3);
    high=matrix(4);
    assoc=matrix(5);
    bach=matrix(6);
    mast=matrix(7);
    phd=matrix(8);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
a=zeros(1,8);                   %Creates array of zeros with specified
    size
[rows, columns]=size(a);


a(1)=1;                         %Represents the individual who will
                                %move through the levels of educ
                                %based on their assigned values
if rand < elem
    a(1)=a(1)-1;
    a(2)=a(2)+1;
    if rand < mid
        a(2)=a(2)-1;
        a(3)=a(3)+1;
        if rand < high
            a(3)=a(3)-1;
            a(4)=a(4)+1;
            if rand < assoc
                a(4)=a(4)-1;
                a(5)=a(5)+1;
                if rand < bach
                    a(5)=a(5)-1;
                    a(6)=a(6)+1;
                    if rand < mast
                        a(6)=a(6)-1;
                        a(7)=a(7)+1;
                        if rand < phd
                            a(7)=a(7)-1;
                            a(8)=a(8)+1;
                        end
                    end
                end
            end
        end
    end
end

if a(1)==1
    outcome = 1; % did not finish any schooling
elseif a(2) ==1
    outcome = 2; % only completed elementary
elseif a(3)==1
    outcome = 3; % only completed middle school
elseif a(4)==1
    outcome = 4; % only completed high school
elseif a(5) ==1
    outcome = 5; % only completed associate's
elseif a(6)==1
    outcome = 6; % only completed bachelor's
elseif a(7)==1
    outcome = 7; % only completed master's
elseif a(8)==1
    outcome = 8; % completed phd
else
    outcome = 0;
end
```

C.4 Generating Data and Running Educational Outcomes

```
% Singular_Model([1 0 1 1 1],1)

% Generating Conditional Probabilities
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
numppl = 200000;

C1 = rand(numppl,1); %Gender
C2 = rand(numppl,1); %Race
C3 = rand(numppl,1); %CVisits
C4 = rand(numppl,1); %InspTeach
C5 = rand(numppl,1); %ParentEd
C6 = rand(numppl,1); %Outcome (0-7)
randmat = horzcat(C1,C2,C3,C4,C5,C6);
sampledata = zeros(numppl,6);
% assigning gender
for j=1:numppl
    if randmat(j,1)<.5
        sampledata(j,1)=0;
    else
        sampledata(j,1)=1;
    end
end
% assigning race
for j=1:numppl
    if randmat(j,2)<.785535
        sampledata(j,2)=0;
    elseif (.785535 <= randmat(j,2))&& (randmat(j,2)<.908575)
        sampledata(j,2)=1;
    else
        sampledata(j,2)=2;
    end
end
% assigning cvisits
for j=1:numppl
    if randmat(j,3)<.240862
        sampledata(j,3)=1;
    else
        sampledata(j,3)=0;
    end
end
% assigning inspteach
for j=1:numppl
    if randmat(j,4)<.124031
        sampledata(j,3)=0;
    else
        sampledata(j,4)=1;
    end
end
% assigning parenteduc
for j=1:numppl
    if randmat(j,5)<.315384615
        sampledata(j,5)=0;
    else
        sampledata(j,5)=1;
    end
end
% assigning outcome
for j=1:numppl
    if randmat(j,6)<.0063
        sampledata(j,6)=0;
    elseif (.0063 <= randmat(j,6))&& (randmat(j,6)<.0203)
        sampledata(j,6)=1;
    elseif (.0203 <= randmat(j,6))&& (randmat(j,6)<.0985)
        sampledata(j,6)=2;
    elseif (.0985 <= randmat(j,6))&& (randmat(j,6)<.5449)
        sampledata(j,6)=3;
    elseif (.5449 <= randmat(j,6))&& (randmat(j,6)<.6467)
        sampledata(j,6)=4;
    elseif (.6467 <= randmat(j,6))&& (randmat(j,6)<.8661)
        sampledata(j,6)=5;
    elseif (.8661 <= randmat(j,6))&& (randmat(j,6)<.9763)
        sampledata(j,6)=6;
    else
        sampledata(j,6)=7;
    end
end

C = unique(sampledata);
sampledataout = [C,histc(sampledata,C)];


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Value  Gender   Race      CVisits   InspTeach   ParentEd
%---------------------------------------------------------------
% 0      Male     White     No        No          First gen
% 1      Female   Black     Yes       Yes         Not first gen
% 2               Other
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
fwyyf_mat(1,6)=0;
fwyyn_mat(1,6)=0;
fwynf_mat(1,6)=0;
fwynn_mat(1,6)=0;
fwnyf_mat(1,6)=0;
fwnyn_mat(1,6)=0;
fwnnf_mat(1,6)=0;
fwnnn_mat(1,6)=0;

fbyyf_mat(1,6)=0;
fbyyn_mat(1,6)=0;
fbynf_mat(1,6)=0;
fbynn_mat(1,6)=0;
fbnyf_mat(1,6)=0;
fbnyn_mat(1,6)=0;
fbnnf_mat(1,6)=0;
fbnnn_mat(1,6)=0;

foyyf_mat(1,6)=0;
foyyn_mat(1,6)=0;
foynf_mat(1,6)=0;
foynn_mat(1,6)=0;
fonyf_mat(1,6)=0;
fonyn_mat(1,6)=0;
fonnf_mat(1,6)=0;
fonnn_mat(1,6)=0;

mwyyf_mat(1,6)=0;
mwyyn_mat(1,6)=0;
mwynf_mat(1,6)=0;
mwynn_mat(1,6)=0;
mwnyf_mat(1,6)=0;
mwnyn_mat(1,6)=0;
mwnnf_mat(1,6)=0;
```

```
mwnnn_mat(1,6)=0;

mbyyf_mat(1,6)=0;
mbyyn_mat(1,6)=0;
mbynf_mat(1,6)=0;
mbynn_mat(1,6)=0;
mbnyf_mat(1,6)=0;
mbnyn_mat(1,6)=0;
mbnnf_mat(1,6)=0;
mbnnn_mat(1,6)=0;

moyyf_mat(1,6)=0;
moyyn_mat(1,6)=0;
moynf_mat(1,6)=0;
moynn_mat(1,6)=0;
monyf_mat(1,6)=0;
monyn_mat(1,6)=0;
monnf_mat(1,6)=0;
monnn_mat(1,6)=0;

for i=1:numppl
    if sampledata(i,1:5)==[0 0 0 0 0]
        mwnnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 0 0 1]
        mwnnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 0 1 0]
        mwnyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 0 1 1]
        mwnyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 1 0 0]
        mwynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 1 0 1]
        mwynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 1 1 0]
        mwyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 0 1 1 1]
        mwyyn_mat(end+1,:)=sampledata(i,:);

    elseif sampledata(i,1:5)==[0 1 0 0 0]
        mbnnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 0 0 1]
        mbnnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 0 1 0]
        mbnyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 0 1 1]
        mbnyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 1 0 0]
        mbynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 1 0 1]
        mbynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 1 1 0]
        mbyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 1 1 1 1]
        mbyyn_mat(end+1,:)=sampledata(i,:);

    elseif sampledata(i,1:5)==[0 2 0 0 0]
        monnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 0 0 1]
        monnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 0 1 0]
        monyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 0 1 1]
        monyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 1 0 0]
        moynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 1 0 1]
        moynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 1 1 0]
        moyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[0 2 1 1 1]
        moyyn_mat(end+1,:)=sampledata(i,:);


    elseif sampledata(i,1:5)==[1 0 0 0 0]
        fwnnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 0 0 1]
        fwnnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 0 1 0]
        fwnyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 0 1 1]
        fwnyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 1 0 0]
        fwynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 1 0 1]
        fwynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 1 1 0]
        fwyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 0 1 1 1]
        fwyyn_mat(end+1,:)=sampledata(i,:);

    elseif sampledata(i,1:5)==[1 1 0 0 0]
        fbnnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 0 0 1]
        fbnnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 0 1 0]
        fbnyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 0 1 1]
        fbnyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 1 0 0]
        fbynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 1 0 1]
        fbynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 1 1 0]
        fbyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 1 1 1 1]
        fbyyn_mat(end+1,:)=sampledata(i,:);

    elseif sampledata(i,1:5)==[1 2 0 0 0]
        fonnf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 0 0 1]
        fonnn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 0 1 0]
        fonyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 0 1 1]
        fonyn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 1 0 0]
        foynf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 1 0 1]
        foynn_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 1 1 0]
        foyyf_mat(end+1,:)=sampledata(i,:);
    elseif sampledata(i,1:5)==[1 2 1 1 1]
        foyyn_mat(end+1,:)=sampledata(i,:);

    end
end
count=[unique(fwyyf_mat),histc(fwyyf_mat,unique(fwyyf_mat))];
pct=count(:,7)./length(fwyyf_mat);
fwyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
        pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
```

```
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fwyyn_mat),histc(fwyyn_mat,unique(fwyyn_mat))];
if length(count)>2
pct=count(:,7)./length(fwyyn_mat);
fwyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
end
count=[unique(fwynf_mat),histc(fwynf_mat,unique(fwynf_mat))];
if length(count)>2
    pct=count(:,7)./length(fwynf_mat);
    fwynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
    fwynf_cum_pct=[1
        .9909
        .9775
        .9079
        .4224
        .3553
        .1324
        .0156];
end
count=[unique(fwynn_mat),histc(fwynn_mat,unique(fwynn_mat))];
if length(count)>2
pct=count(:,7)./length(fwynn_mat);
fwynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
    fwynn_cum_pct=[1
        .9909
        .9775
        .9079
        .4224
        .3553
        .1324
        .0156];
end
count=[unique(fwnyf_mat),histc(fwnyf_mat,unique(fwnyf_mat))];
pct=count(:,7)./length(fwnyf_mat);
fwnyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fwnyn_mat),histc(fwnyn_mat,unique(fwnyn_mat))];
pct=count(:,7)./length(fwnyn_mat);
fwnyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fwnnf_mat),histc(fwnnf_mat,unique(fwnnf_mat))];
pct=count(:,7)./length(fwnnf_mat);
fwnnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fwnnn_mat),histc(fwnnn_mat,unique(fwnnn_mat))];
pct=count(:,7)./length(fwnnn_mat);
fwnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fwnnn_mat),histc(fwnnn_mat,unique(fwnnn_mat))];
pct=count(:,7)./length(fwnnn_mat);
fwnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fbyyf_mat),histc(fbyyf_mat,unique(fbyyf_mat))];
pct=count(:,7)./length(fbyyf_mat);
fbyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(fbyyn_mat),histc(fbyyn_mat,unique(fbyyn_mat))];
pct=count(:,7)./length(fbyyn_mat);
fbyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
```

```
          pct(8)];
count=[unique(fbynf_mat),histc(fbynf_mat,unique(fbynf_mat))];
if length(count)>2
pct=count(:,7)./length(fbynf_mat);
fbynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
else
    fbynf_cum_pct=[1
         .9919
         .9852
         .8812
         .3377
         .2691
         .1001
         .0126];
end
count=[unique(fbynn_mat),histc(fbynn_mat,unique(fbynn_mat))];
if length(count)>2
pct=count(:,7)./length(fbynn_mat);
fbynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
else
    fbynn_cum_pct=[1
         .9919
         .9852
         .8812
         .3377
         .2691
         .1001
         .0126];
end
count=[unique(fbnyf_mat),histc(fbnyf_mat,unique(fbnyf_mat))];
pct=count(:,7)./length(fbnyf_mat);
fbnyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(fbnyn_mat),histc(fbnyn_mat,unique(fbnyn_mat))];
pct=count(:,7)./length(fbnyn_mat);
fbnyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(fbnnf_mat),histc(fbnnf_mat,unique(fbnnf_mat))];
pct=count(:,7)./length(fbnnf_mat);
fbnnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(fbnnn_mat),histc(fbnnn_mat,unique(fbnnn_mat))];
pct=count(:,7)./length(fbnnn_mat);
fbnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(foyyf_mat),histc(foyyf_mat,unique(foyyf_mat))];
pct=count(:,7)./length(foyyf_mat);
foyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(foyyn_mat),histc(foyyn_mat,unique(foyyn_mat))];
pct=count(:,7)./length(foyyn_mat);
foyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
count=[unique(foynf_mat),histc(foynf_mat,unique(foynf_mat))];
if length(count)>2
pct=count(:,7)./length(foynf_mat);
foynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(5)+pct(6)+pct(7)+pct(8)
          pct(6)+pct(7)+pct(8)
          pct(7)+pct(8)
          pct(8)];
else
    foynf_cum_pct=[1
         .9946
         .9910
         .8559
         .2230
         .1887
         .0253
         .0006];
end
count=[unique(foynn_mat),histc(foynn_mat,unique(foynn_mat))];
if length(count)>2
pct=count(:,7)./length(foynn_mat);
foynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
          pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
```

```
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
else
    foynn_cum_pct=[1
            .9946
            .9910
            .8559
            .2230
            .1887
            .0253
            .0006];
end
count=[unique(fonyf_mat),histc(fonyf_mat,unique(fonyf_mat))];
pct=count(:,7)./length(fonyf_mat);
fonyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(fonyn_mat),histc(fonyn_mat,unique(fonyn_mat))];
pct=count(:,7)./length(fonyn_mat);
fonyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(fonnf_mat),histc(fonnf_mat,unique(fonnf_mat))];
pct=count(:,7)./length(fonnf_mat);
fonnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(fonnn_mat),histc(fonnn_mat,unique(fonnn_mat))];
pct=count(:,7)./length(fonnn_mat);
fonnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(mwyyf_mat),histc(mwyyf_mat,unique(mwyyf_mat))];
pct=count(:,7)./length(mwyyf_mat);
mwyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(mwyyn_mat),histc(mwyyn_mat,unique(mwyyn_mat))];
pct=count(:,7)./length(mwyyn_mat);
mwyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(mwynf_mat),histc(mwynf_mat,unique(mwynf_mat))];
if length(count)>2
pct=count(:,7)./length(mwynf_mat);
mwynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
else
    mwynf_cum_pct=[1
            .9910
            .9785
            .8964
            .3987
            .3490
            .1267
            .0239];
end
count=[unique(mwynn_mat),histc(mwynn_mat,unique(mwynn_mat))];
if length(count)>2
pct=count(:,7)./length(mwynn_mat);
mwynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
else
    mwynn_cum_pct=[1
            .9910
            .9785
            .8964
            .3987
            .3490
            .1267
            .0239];

end
count=[unique(mwnyf_mat),histc(mwnyf_mat,unique(mwnyf_mat))];
pct=count(:,7)./length(mwnyf_mat);
mwnyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
              pct(5)+pct(6)+pct(7)+pct(8)
              pct(6)+pct(7)+pct(8)
              pct(7)+pct(8)
              pct(8)];
count=[unique(mwnyn_mat),histc(mwnyn_mat,unique(mwnyn_mat))];
pct=count(:,7)./length(mwnyn_mat);
```

```
mwnyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mwnnf_mat),histc(mwnnf_mat,unique(mwnnf_mat))];
pct=count(:,7)./length(mwnnf_mat);
mwnnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mwnnn_mat),histc(mwnnn_mat,unique(mwnnn_mat))];
pct=count(:,7)./length(mwnnn_mat);
mwnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mwnnn_mat),histc(mwnnn_mat,unique(mwnnn_mat))];
pct=count(:,7)./length(mwnnn_mat);
mwnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mbyyf_mat),histc(mbyyf_mat,unique(mbyyf_mat))];
pct=count(:,7)./length(mbyyf_mat);
mbyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mbyyn_mat),histc(mbyyn_mat,unique(mbyyn_mat))];
pct=count(:,7)./length(mbyyn_mat);
mbyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mbynf_mat),histc(mbynf_mat,unique(mbynf_mat))];
if length(count)>2
pct=count(:,7)./length(mbynf_mat);
mbynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
    mbynf_cum_pct=[1
        .9929
        .9859
        .8773
        .2830
        .2317
        .0754
        .0114];
end
count=[unique(mbynn_mat),histc(mbynn_mat,unique(mbynn_mat))];
if length(count)>2
pct=count(:,7)./length(mbynn_mat);
mbynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
    mbynn_cum_pct=[1
        .9929
        .9859
        .8773
        .2830
        .2317
        .0754
        .0114];
end
count=[unique(mbnyf_mat),histc(mbnyf_mat,unique(mbnyf_mat))];
pct=count(:,7)./length(mbnyf_mat);
mbnyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mbnyn_mat),histc(mbnyn_mat,unique(mbnyn_mat))];
pct=count(:,7)./length(mbnyn_mat);
mbnyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(mbnnf_mat),histc(mbnnf_mat,unique(mbnnf_mat))];
pct=count(:,7)./length(mbnnf_mat);
mbnnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
```

```
count=[unique(mbnnn_mat),histc(mbnnn_mat,unique(mbnnn_mat))];
pct=count(:,7)./length(mbnnn_mat);
mbnnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(moyyf_mat),histc(moyyf_mat,unique(moyyf_mat))];
pct=count(:,7)./length(moyyf_mat);
moyyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(moyyn_mat),histc(moyyn_mat,unique(moyyn_mat))];
pct=count(:,7)./length(moyyn_mat);
moyyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(moynf_mat),histc(moynf_mat,unique(moynf_mat))];
if length(count)>2
     pct=count(:,7)./length(moynf_mat);
moynf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
     moynf_cum_pct=[1
          .9994
          .9977
          .8515
          .1696
          .1450
          .0241
          .0053];
end
count=[unique(moynn_mat),histc(moynn_mat,unique(moynn_mat))];
if length(count)>2
pct=count(:,7)./length(moynn_mat);
moynn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
else
     moynn_cum_pct=[1
          .9994
          .9977
          .8515
          .1696
          .1450
          .0241
          .0053];
end
count=[unique(monyf_mat),histc(monyf_mat,unique(monyf_mat))];
pct=count(:,7)./length(monyf_mat);
monyf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(monyn_mat),histc(monyn_mat,unique(monyn_mat))];
pct=count(:,7)./length(monyn_mat);
monyn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(monnf_mat),histc(monnf_mat,unique(monnf_mat))];
pct=count(:,7)./length(monnf_mat);
monnf_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
count=[unique(monnn_mat),histc(monnn_mat,unique(monnn_mat))];
pct=count(:,7)./length(monnn_mat);
monnn_cum_pct=[pct(1)+pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(2)+pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(3)+pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(4)+pct(5)+pct(6)+pct(7)+pct(8)
            pct(5)+pct(6)+pct(7)+pct(8)
            pct(6)+pct(7)+pct(8)
            pct(7)+pct(8)
            pct(8)];
cum_pct_matrices = {
mwnnf_cum_pct
mwnnn_cum_pct
mwnyf_cum_pct
mwnyn_cum_pct
mwynf_cum_pct
mwynn_cum_pct
mwyyf_cum_pct
mwyyn_cum_pct
mbnnf_cum_pct
mbnnn_cum_pct
mbnyf_cum_pct
mbnyn_cum_pct
mbynf_cum_pct
mbynn_cum_pct
mbyyf_cum_pct
mbyyn_cum_pct
monnf_cum_pct
```

```
monnn_cum_pct
monyf_cum_pct
monyn_cum_pct
moynf_cum_pct
moynn_cum_pct
moyyf_cum_pct
moyyn_cum_pct
fwnnf_cum_pct
fwnnn_cum_pct
fwnyf_cum_pct
fwnyn_cum_pct
fwynf_cum_pct
fwynn_cum_pct
fwyyf_cum_pct
fwyyn_cum_pct
fbnnf_cum_pct
fbnnn_cum_pct
fbnyf_cum_pct
fbnyn_cum_pct
fbynf_cum_pct
fbynn_cum_pct
fbyyf_cum_pct
fbyyn_cum_pct
fonnf_cum_pct
fonnn_cum_pct
fonyf_cum_pct
fonyn_cum_pct
foynf_cum_pct
foynn_cum_pct
foyyf_cum_pct
foyyn_cum_pct
    };
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Value   Gender    Race        CVisits    InspTeach    ParentEd
%-----------------------------------------------------------------------
% 0         Male      White       No         No           First gen
% 1         Female    Black       Yes        Yes          One parent
% 2                   Other                               Both parent
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
numppl2 = 10000;
C1 = randi([0 1],[numppl2 1]);
C2 = randi([0 2],[numppl2 1]);
C3 = randi([0 1],[numppl2 1]);
C4 = randi([0 1],[numppl2 1]);
C5 = randi([0 1],[numppl2 1]);
People = horzcat(C1,C2,C3,C4,C5);
[R, C]=size(People);
outcomeMatrix = zeros(R,1);
for i=1:R
    outcomeMatrix(i) = Singular_Model_3(People,i,cum_pct_matrices);
end
B = unique(outcomeMatrix);
out = [B,histc(outcomeMatrix,B)];
bar(out(:,2))
set(gca,'xticklabel',{out(:,1)})
barvalues;
```