Teacher Population Final Report

Michael Goodenbour, Srujana Dandem, Towera Longwe

May 25, 2023

1 Executive Summary

Teacher shortage has been a prominent issue in the workforce for as long as we can remember. However, from our very small view of the world, we have not seen this issue first hand. Especially being education majors, we feel like there are plenty of future educators in our classes and out in the world. So, this provided our motivation to model this problem. Our overall question that we looked at is: How has the population of teachers changed over time? We broke that question down into three different questions with three models in hopes of modeling and predicting the teacher population in the future.

Our paper takes a novel approach by looking at modeling the number of teachers as a whole at a given point in time. Using our knowledge of differential equations, we were able to find the rate at which teachers are leaving the profession, helping to model future teacher populations. We also wanted to examine what factors most affect the number of teachers in a given year. We looked at four factors: yearly salary, economy (indicated by GDP per capita), student-teacher ratio, and total population in the United States. After using a little linear algebra, we were able to find values for our unknown coefficients. After conducting a sensitivity analysis, our model showed that total population in the United States most influences the number of teachers in a given year in the United States. We also wanted to estimate the number of teachers for given areas of study to best predict which areas of study might have shortages in the future. We created a model that incorporated the graduation numbers of future teachers in that area of study as well as the retirement rate of teachers in that area of study through simple difference equations.

After crunching the data, both of our working models concluded that the population of teachers will not decline any time soon. Our first model concluded that the number of teachers will slowly increase before remaining constant. However, we predict that it will take around another 140 years or so for this to happen. Our second model concluded that the number of teachers will continue to increase as long as the current trends in our factors continue. However, if one of those factors were to change, the population in teachers could rapidly decrease. As a whole, our models do not seem to spit out any data that would suggest a drastic teacher shortage in the upcoming years.

2 Question 1

2.1 Introduction

The first question we decided to look at was: what is the overall trend in the population of teachers in the United States?

We wanted to make a simple model that just looked at the number of incoming teachers and the rate at which teachers are leaving to model the total population of teachers.

2.2 Model

2.2.1 Assumptions

We established a few assumptions for our first model. The first assumption that we made for our model is that we only looked at teachers in the United States. We chose to narrow it to just teachers in the United States, because there was very little data on teachers worldwide. Data about teachers in the United States was much more plentiful.

Another assumption we made was that the number of incoming teachers yearly was fixed (does not vary from year to year). We decided to make this a fixed rate, because having a fluctuating number would make our model much more complex, something that we did not have the time for.

The last assumption we made was that the rate at which teachers were leaving the profession was also fixed. By assuming this, it allowed us to create a simple differential equation for the change of teachers annually.

2.2.2 Variables

We decided to include a couple different variables in our model. Those variables are listed below in Table 1, along with their labels.

2.2.3 The Model

Using our variables, we were able to make a simple differential equation that produced the rate of change in the population of teachers annually.

$$\frac{dT}{dt} = g - rT$$

However, we wanted to model the total population of teachers, rather than the rate at which that population changes. So, we integrated that equation to produce a model that gave us the teacher population at any given year:

$$T = k - Ae^{-rt}$$

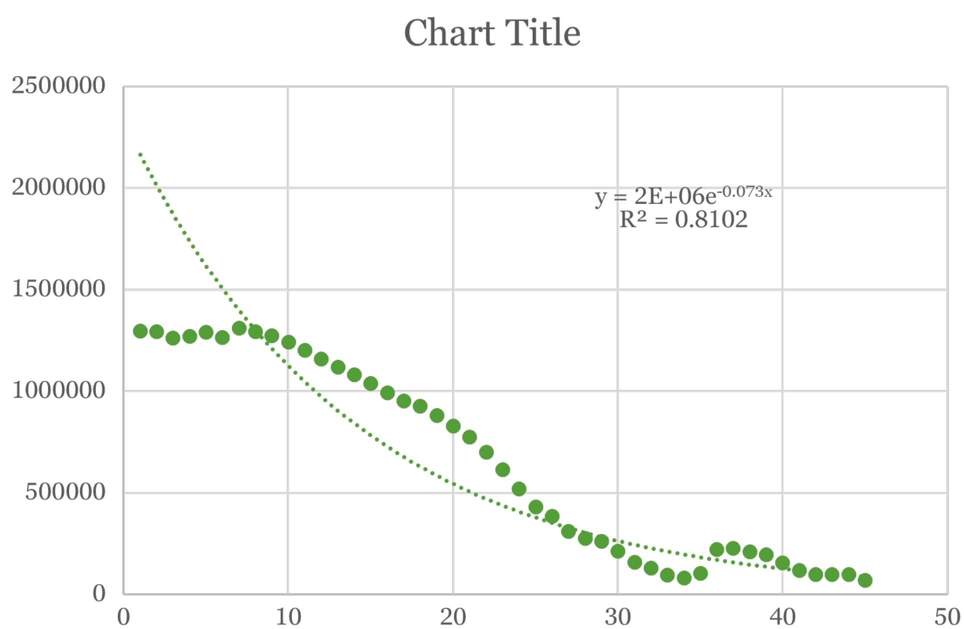
However, we didn't know what our actual r value was. So, we decided to use Excel to model this for us.

2.2.4 Model Explanation

First, we researched that the average teacher turnover rate (rate at which teachers leave the profession) had been hovering at around 8% for the past couple decades[1]. We used that predicted teacher turnover rate ($r_p = 0.08$), in order to determine the r value for our predicted population. In order to do, we graphed t vs $k - T$. From our previous integration, we know that our k coefficient is actually the carrying capacity [2] of our teachers. The carrying capacity can otherwise be written has $\frac{g}{r}$, or $\frac{300{,}000}{0.08} = 375,000$ teachers.

When graphing t vs $k - T$ in Excel, we got an equation of:

$$k - T = 2000000e^{-0.073t}$$



Graph 1: Graph of actual teacher population (in dots) vs. predicted teacher population (line)

We were able to find that our actual teacher turnover rate ended up being 7.3%, which is pretty similar to our predicted 8%.

[1] https://www.mckinsey.com/industries/education/our-insights/k-12-teachers-are-quitting-what-would-make-them-stay

[2] As Dr. Birgen mentioned, we understand that carrying capacity has a specific definition that is only associated with logistic models. However, instead of changing this, we decided to continue using the phrase carrying capacity. Just note that it is being used outside of its typical definition.

2.3 Solution and Discussion

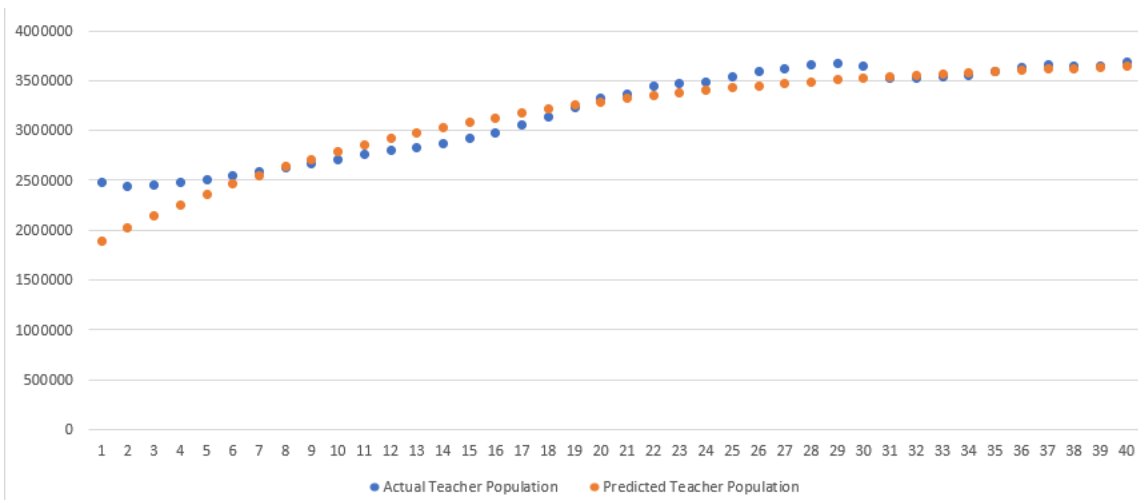
2.3.1 Solution

Using the values that we found for our variables and coefficients, we are able to model the teacher population at any given year through our model:

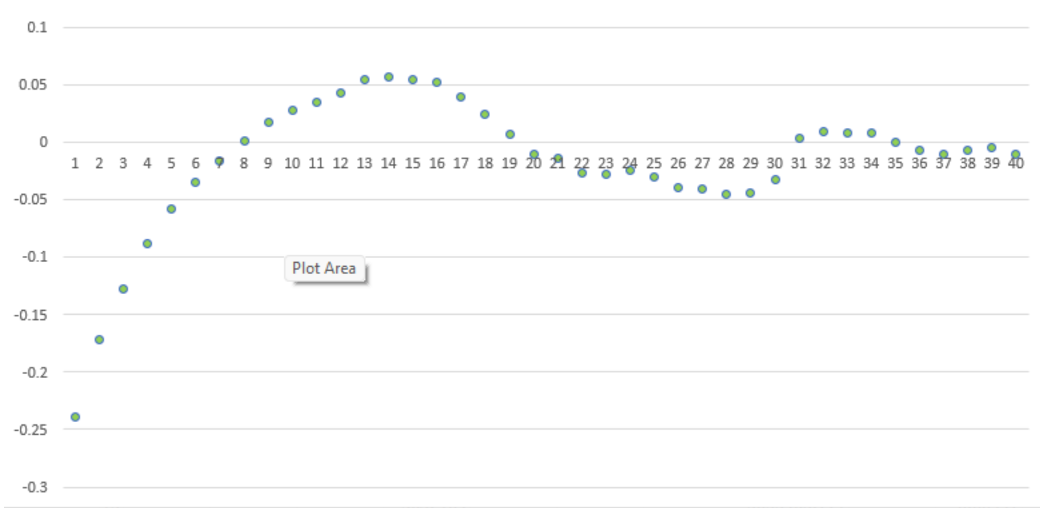
$$T = 375,000 - 2,000,000e^{-0.073t}$$

Variable	Quantity	Units
g	300,000	teachers [3]
r	0.073	unitless
k	375,000	teachers

Table 4: List of found values for variables used in model



Graph 2: Actual Teacher Population vs. Our Predicted Teacher Population for Model 1



Graph 3: Graph of Residuals for Model 1

2.3.2 Implications

Our model revealed that the teacher population will continue to slowly increase until its reaches its carrying capacity. We can conclude that according to our model, teacher shortage will not be an issue in upcoming years. Teacher population will not reach its carrying capacity until the year 2184, and until then, the population will continuously increase at a slower rate.

2.3.3 Validity of the Model

When examining the validity of this model, we looked at the different parts of our model that made logical sense and looked at the different parts that didn't make any logical sense at all.

Naturally, the world has a carrying capacity. There are only a certain amount of people that this world can provide for. Therefore, this makes sense for there to also a carrying capacity for the population of teachers. If the population of the United States is no longer increasing, the need for more teachers would also become stagnant.

However, technology has the ability to change this carrying capacity. The technology we have developed to produce more food, provide more living space, etc. has already changed the carrying capacity in this world. We do not know if the carrying capacity of the world right now will be in the same as it will be in the future, so this would change the carrying capacity of our teachers. However, that is an nearly impossible problem to model.

2.3.4 Strengths and Weaknesses

As far a strengths of our model, we were able to determine what we believe to be the carrying capacity of teachers. The number of teachers should not exceed this threshold, as there will not be enough need for them as the population in the United States also becomes stagnant. Another strength of our model is its accuracy. Despite some outliers in the beginning few years of data, our predictions were no more than around 6% off from the actual teacher population, which means that we should be able to accurately predict the teacher population in the future.

However, our model obviously had quite a few weaknesses as well. Our model simply used the trend of past data to predict the future population of teachers. We did not account for any of the factors that influence the population of teachers. These factors could cause the population to vary from the trend in past years. For example, if the government suddenly decided to pay teachers significantly less money, there would probably be a decrease in the teacher population, as teachers would leave and look for other jobs. As we mentioned above, the carrying capacity could also vary as technology improves, so the carrying capacity of teachers would vary from our projected number. Our model was also very poor at accurately modeling the teacher population during the first few years. As the years become more recent, our predictions become more accurate, but our final prediction in 1975 was almost 25% off. Starting around 1981, our model is no more than around 6% off, so we could just advice people to use our model for years 1981 and later.

2.4 Sensitivity Analysis

Using the formula for our model, $T = k - Ae^{-rt}$, we decided to increase each variables by five percent to see how it impacts the total teacher population.

Variable	Change	Change of T
k	5% increase	6.11% increase in total population
r	5% increase	0.82% increase in total population

Table 5: Sensitivity Analysis of Model 1

As you can see, our model is not that sensitive, as a 5% increase in each variable does not lead to a significantly larger chance in the total teacher population. Our model is most influenced by k, our carrying capacity.

3 Question 2

3.1 Introduction

The second question we decided to look at was: what factors influence the population of teachers and how do they impact it? We wanted to look at a slightly more complex model that showed the population of teachers in the United States at any given past year by using four factors: salary, economy, student/teacher ratio, and total population of the United States. In order to form our model, we looked at date of those four factors from the years 1980-2019.

3.2 Model

3.2.1 Assumptions

We only had a couple assumptions for our second model. The first assumption is the same at the first model in that we only looked at teachers in the United States. Our reasoning can be found in the first model assumption section. [4]

Another assumption we had for this model is that the population is solely affected by the four factors that we looked at. Obviously, the population got constricted to solely these four factors, but we decided on these four factors because they were the factors that we brainstormed that also had substantial data over time.

3.2.2 Variables

We decided to include four different variables and four different coefficients for our second model. Those variables are listed below in Table 2, along with their labels.

3.2.3 The Model

Using our four variables and four coefficients, we were able to make a simple linear model that implementing our four factors that influence teacher population to predict the total population of teachers.

$$T = as + be + cs_{tr} + dp$$

3.2.4 Model Explanation

To first start the modeling process, we first researched data for each of our four factors that influence the teacher population. We found data for the average salary, GDP per capita in the United States, student/teacher ratio, and total population of the United States for every year from 1980 to 2019.

At first, we decided to just look at three different factors, excluded the total United States population factor. In order to solve for three unknown coefficients (a,b,c), we first decided to use a system of equations using the data from three random years and putting that data into a matrix. We produced a 3x4 matrix (Table 3), with the columns being average salary, GDP per capita, student/teacher ratio, and total teacher population. We used SAGE to row reduce our matrix, giving us our coefficient values in the fourth column.

58016	44661.11	18.6	2485216
69647	61743.00	15.9	3365867
70315	74544.44	15.3	3679371

[4] We chose to narrow it to just teachers in the United States, because there was very little data on teachers worldwide. Data about teachers in the United States was much more plentiful.

Table 6: Original Matrix for our System of Equations

1	0	0	37.225
0	1	0	21.004
0	0	1	-32927.95

Table 7: Row Reduced Matrix for our System of Equations

Therefore, our model using three random years of data is $37.225a + 21.004b - 32927.95c$ and we thought we were done for this part.

However, Dr. Mr. Birgen came in and gave us some advice on how to improve this model. When using a system of equations, it gave us the exact answers of the coefficients for those three equations. However, the reason why this method is not a good representative of all forty years of data is because one of the random years chosen could be in outlier. For example, the economy could be abnormally poor that random year, and that would significantly skew our model, causing our model to be a poor representative of all of our data as a whole.

Instead of using a random three years, we decided to use all forty years of data, as well as added our fourth factor, the total population of the United States. We learned from linear algebra that $A^T Ax = A^T b$. A is our 40x4 matrix of all of our data, with the columns being our different factors, and the rows being every year of data. b is a 40x1 vector of the population of teachers every year since 1980. x is a 4x1 vector of the unknown coefficients (a,b,c,d) that we are trying to solve.

When solving for x, we first found that AA^T was a 4x4 matrix (Image 1) and $A^T b$ was a 4x1 vector (Image 2).

```
[7.89188225390000e10 6.93761050525886e10  2.72373127000000e7 4.89402215295236e14]
[6.93761050525886e10 6.16393950105493e10  2.31560651732750e7 4.23105043527806e14]
[ 2.72373127000000e7  2.31560651732750e7    10794.8900000000 1.81534509910000e11]
[4.89402215295236e14 4.23105043527806e14 1.81534509910000e11 3.14984156734640e18]
```

Image 1: AA^T matrix

```
[ 5.60666732200000e9]
[ 4.85256593019005e9]
[ 2.06156120000000e6]
[3.59322786320920e13]
```

Image 2: ATb vector

We let the product of $AA^T = M$ and the product of $A^T b = N$ just for easier clarity. Using substitution, we know that $Mx = N$. From linear algebra, in order to solve for x, we must take $M^{-1} * N = x$. Computing that, we get a 4x1 vector (Image 3) with our solutions for our four different coefficients.

```
[  -0.0332793079261933]
[  -0.0246984559004639]
[    -216.202505873633]
[0.0000323563902118323]
```

Image 3: x (solution) vector

3.3 Solution and Discussion

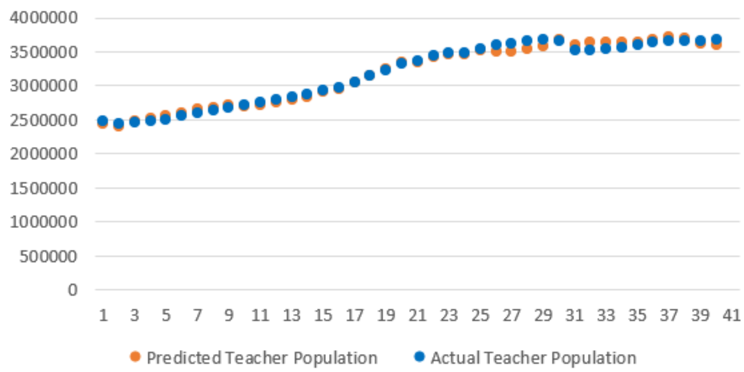
3.3.1 Solution

Using the values that we solved for in the x vector, we are able to model the teacher population in past years based through those four different factors:

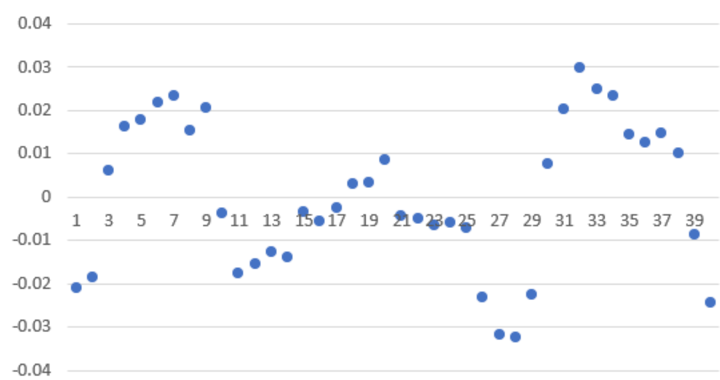
$$T = -0.0333a - 0.0247b - 216.203s_{tr} + 0.000032p$$

Variable	Number	Units
a	-0.0333	teachers [5]
b	-0.0247	teachers [6]
c	-216.203	teachers [5]
d	0.000032	teachers [7]

Table 8: List of found coefficients used in model



Graph 4: Actual Teacher Population vs. Our Predicted Teacher Population for Model 2



Graph 5: Residual Plot for Model 2

3.3.2 Implications

Our model revealed that the teacher population will continue to slowly increase as long as the past trends in our four factors continue into the future. Therefore, we conclude that according to our model, teacher shortage will not be an issue in the upcoming years. One of our factors, such as a huge increase in student/teacher ratio or a huge decrease the population of the United States, would have to occur into our for the population of our teachers to be in jeopardy.

3.3.3 Validity of the Model

When looking for validity of the model, we looked at the relationship between our different factors and the population of teachers. Two of our relationships made logical sense, and two did not.

The two factors that made sense were the student/teacher ratio and the population of the United States. The coefficient of the student/teacher ratio (c) was negative, meaning that there is a inverse relationship. This makes sense because as the student/teacher ratio increases (more students for less teachers), it makes sense that there would be less teachers available to teach the same amount of kids. The other factors that made sense was the total population of the United States. The coefficient of the total population of the United States (d) was positive, meaning there is a positive relationship. This makes sense because as the population of the United States increases, this means there are going to be more children to teach, hence the need and growth of more teachers.

The two factors that did not make sense were the economy and salary. Both coefficients were negative, meaning there is an inverse relationship. This doesn't make sense, because as the economy and salary both improve, we should see more teachers wanting to go into the profession. However, according to our model, the opposite happens, and more teachers leave the profession. We believe this is the case because both of these variables are correlated, which causes our model to be over-described.

3.3.4 Strengths and Weaknesses

As far as strengths of our model, we were able to model the teacher population based our four different factors, rather than just using a fixed number of graduates and turnover rate. Another strength of our model is that it is relatively easy to calculate. Someone simply needs to just research data on a specific year for our four different values, and plug them into the variables, and our model will tell you the predicted number of teachers nationwide. This model was also incredibly accurate. Our model was at most 3% off at its most inaccurate points, and a lot of data hovering inside the 2% error range. This model was much more accurate than our first model.

However, our model had some weaknesses as well. The model looks at solely our four factors, and disregards all other factors, such as how the development of technology has impacted the numbers of teachers. There are endless different factors that influence teacher population, but we were only able to look at these four. This model is also hard to use when trying to predict a population. Since this model uses data based on past years, someone would have to use future predictions for those values, which may not be super accurate. For example, the projection is teacher salary could vary drastically if a new law is passed that cut down of government funding for the education department.

3.4 Sensitivity Analysis

Using the formula for our model, $T = -0.0333a - 0.0247b - 216.203s_{tr} + 0.000032p$, we decided to increase each variables by five percent to see how it impacts the total teacher population. As you can see below, the change in the total population of the United States was the most sensitive of our variables, which makes sense as the population is very large, and a 5% increase would be substantial. Overall, our model is not that sensitive.

Variable	Change	Change of T
s	5% increase	2.16% increase in total population
e	5% increase	1.36% increase in total population
s_{tr}	5% increase	5.76% increase in total population
p	5% increase	14.28% increase in total population
a	5% increase	2.16% increase in total population
b	5% increase	1.36% increase in total population
c	5% increase	5.76% increase in total population
d	5% increase	14.28% increase in total population

Table 9: Sensitivity Analysis of Model 1

4 Question 3

4.1 Introduction

The third question we decided to answer is: what are the areas of study that we won't have enough teachers for in the future? For this question, we decided to look at modeling the population of teachers in different areas of study. We selected Math, History, English and Science as the areas that we would want to model the populations for.

4.2 Model

4.2.1 Assumptions

We had three assumptions for our third model. The first assumption is the same as the other two models in that we are looking at teaching solely in the United States. [8]

Another assumption we had was that all college graduates with education degrees go into teaching that specific subject during their first year out of college.

We assumed this so it would easier to model to assume all graduates go to teach. We know this is unrealistic as people get offered better opportunities and pursue those.

Our last assumption is that we only considered math, science, history, and English teachers. Obviously, those are the four most prominent subjects taught in the United States, but there are other subjects that we could consider if we had more time.

4.2.2 Variables

We decided to use sixteen different variables for this model. However, we developed four different models for the teacher population of each subject we looked at, so it is not as overwhelming as that may seem. Our variables are listed in our Appendices (Table 3).

4.2.3 The Model

Using our sixteen different variables for our four different models, we were able to model a four different simple difference equations.

$$T_{m+1} = T_m + (g_m + r_{lm}(T_m))$$
$$T_{s+1} = T_s + (g_s + r_{ls}(T_s))$$
$$T_{h+1} = T_h + (g_h + r_{lh}(T_h))$$
$$T_{e+1} = T_e + (g_e + r_{le}(T_e))$$

4.2.4 Model Explanation

At this point in the project, we can thrown a majority of our time and effort into the first two parts of the projects, so we wanted to develop more of a simple model to answer an interesting question. We decided on just developing four different simple difference equations, where we modeled the future teacher population by adding the present value to the change in the teacher population.

We tried to research data for our unknown variables but did not have any luck, as our time was running short. But, without looking at any data, it is obvious that if the number of teachers are greater than those graduating and coming into the workforce, that specific subject will eventually not enough teachers. Using data from previous years, it would be easy to predict how soon this problem would occur or how quickly the teacher shortage could spiral out of control.

[8]We chose to narrow it to just teachers in the United States, because there was very little data on teachers worldwide. Data about teachers in the United States was much more plentiful.

4.3 Discussion

4.3.1 Potential Problems to Look At

Since we did not find any data, we brainstormed potential different interesting problems we could look at which these models. Obviously, these would allow us to look at which subjects are going to need more teachers in the future. This would allow high schools and college to promote these positions in hopes that the future generations can fill those holes.

I think the coolest thing that we can look at with these models are why different subject areas are failing to produce teachers while others are doing just fine. We could look at the simple factors, such as graduation numbers are too low or the rate of teachers leaving is too high, but we can also look at the opportunities that young graduates are offered straight out of college. Are there some degrees that are offered better non-teaching opportunities than others? There are numerous different directions to go with this model, if we simply had the time.

4.3.2 Strengths and Weaknesses

Some of the strengths of our model are that is it super easy to calculate (if you have the data). Our model is also extremely easy to understand, as we are just simply finding the future value of the teacher population by adding current population to the change in the population. We also don't assume that the rate of teachers leaving and the number of graduates are constant, which is the case in our first model.

Our model also has its weaknesses. This model, similarly to our first model, simply shows the numbers of teachers, not the reasons that influence the population. If we somehow were able to incorporate different "push or pull" [9] factors, we could have a better prediction of the teacher population, as well as solution on how to increase the teacher population in areas of study that are struggling.

5 Conclusions

Our overall goal was to model how the teacher population in the United States has changed over time. We hear often of the increase in teacher shortages in recent years, and as educators, this seems relevant to us.

In the first model, we use trends of the teacher population to predict the future number of teachers. We were also able to find a "carrying" capacity for the population of teachers. We found that the population of teachers would increase slowly until reaching this "carrying" capacity.

In the second model, we looked at four different factors that influence how many teachers there are in the United States in any given year. The factors were yearly salary, economy(as indicated by GDP per capita), student-teacher ratio and the population of the United States. We used linear algebra to find

[9] appealing factors of teaching vs. unappealing factors of teaching

coefficients for each of these factors. We found that the teacher population should slowly increase as long as the recent trends in each of the four factors continue.

In the third model, we looked at different teacher populations based on area of study. We decided to look at four areas of study: Math, English, History and Science. We wanted to see trends in the population to predict which subjects will not have enough teachers in the future. Our model incorporated the number of college graduates going to be teachers in a particular field and turnover rate of current teachers in a particular field. Relevant data for graduation numbers and turnover rates for each area of study was hard to find. However, if we were able to find that data, we could look at reasons for why different teacher populations of different subjects are struggling while others are thriving. Those reasons could include better opportunities for those certain degrees right out of college or in early adulthood.

Based on our findings, we would make recommendations to the Iowa State Board of Education to have incentives for teachers who teach areas of study where the teacher population is struggling. Some degrees, such as mathematics or physics, may have better job opportunities like engineering, that provide better pay opportunities. Increasing the salary of those positions might be one way to help the teacher population of those subjects. Modeling the number of teachers could also help the Iowa State Board of Education better plan for training opportunities for teachers, as well as helping to best plan how for new school districts, and renovations of classrooms and buildings in existing school districts.

For as long as we can remember, teacher shortage has been a prominent issue. However, our models show that the teacher population will continue to increase. Will this slow increase be enough to keep up with the rapidly growing school districts? Our models seem to think so, but only time will truly tell. The Board of Education has endless possibilities with this data, all in hopes to improve that educational experience of every single child nationwide.

6 References

U.S. Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/
National Center for Education Statistics. (n.d.). Retrieved from https://nces.ed.gov/
https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=US
https://fred.stlouisfed.org/series/POPTOTUSA647NWDB

7 Appendices

Table 1: Variables

Variable	Label	Dependent/Independent	Unit
T	Number of teachers at a given time	Dependent	people
g	Incoming teachers	Independent	people
r	Rate of turnover	Independent	unitless
t	Years since 1975	Independent	years
k	Carrying capacity	Independent	people

Table 2: Variables

Variable	Label	Dependent/Independent	Unit
a	Salary coefficient	Independent	teachers
b	GDP per capita coefficient	Independent	teachers
c	Student-Teacher Ratio Coefficient	Independent	teachers
d	Population Coefficient	Independent	teachers
T	Number of teachers at a given time	Dependent	people
s	Average teacher salary in the US for a given year	Independent	Dollars
e	Average GDP per capita of the US for a given year	Independent	Dollars
s_{tr}	Average student-teacher ratio in the US for a given year	Independent	unitless
p	Average population in the US for a given year	Independent	people

Table 3: Variables

Variable	Label	Dependent/Independent	Unit
T_{m+1}	Number of Math Teachers in the following year	Dependent	people
T_m	Number of Math Teachers in the current year	Independent	people
g_m	Number of math graduates in the current year	independent	people
r_{lm}	Rate at which Math Teachers leave the profession	Independent	unitless
T_{s+1}	Number of Science Teachers in the following year	Dependent	people
T_s	Number of Science Teachers in the current year	Independent	people
g_s	Number of science graduates in the current year	independent	people
r_{ls}	Rate at which Science Teachers leave the profession	Independent	unitless
T_{h+1}	Number of History Teachers in the following year	Dependent	people
T_h	Number of History Teachers in the current year	Independent	people
g_h	Number of history graduates in the current year	independent	people
r_{lh}	Rate at which History Teachers leave the profession	Independent	unitless
T_{e+1}	Number of English Teachers in the following year	Dependent	people
T_e	Number of English Teachers in the current year	Independent	people
g_e	Number of English graduates in the current year	independent	people
r_{le}	Rate at which English Teachers leave the profession	Independent	unitless