

Collin DeVore  
Data Visualization Project

## THE FRAGMENTATION OF COLLEGE ENROLLMENT DATA

### Introduction

Many of the studies regarding sex and gender in college enrollment data have attempted to explain the shift in the male/female ratio that occurred around the late 1970s and early 1980s. During this time, males began to enroll in smaller numbers than females, leading researchers to wonder what could be occurring to drive the total number of males enrolled in college down and the total number of females enrolled up? While many were examining this phenomenon, they examined aggregate data of the total number of enrollments of men and women in the United States. Unfortunately, few, if any, disaggregated the number of males and females enrolled in order to comparatively examine different states to see if there is any possibility of geographic, political, or economic differences or shocks that could be influencing the data on a more minuscule level. In examining the state data, it should even be possible to see if the aggregated dataset is representative of these smaller datasets.

In 2018, Agger, Meece, and Byun wrote an article called “The Influences of Family and Place on Rural Adolescents’ Educational Aspirations and Post – secondary Enrollment”, in which they analyze the psychological factors that lead to college and educational attainment, and how these factors are brought about by perceptions of the expectations of parents and job opportunities on rural males and females (Agger, Meece, and Byun 2018). The authors find that males have a higher perception of job opportunities, while females have a higher perception of their parents’ expectations (Agger, Meece, and Byun 2018). It is believed that these two perceptions contribute to the gap in rural male and female enrollments (Agger, Meece, and Byun 2018). Another article, written by Conger and Long and titled “Why are Men Falling Behind? Gender Gaps in College Performance and Persistence”, further analyzed the reason that males are enrolling less than females, finding that the issue of males having lower test

scores and difficulty continuing in college starts in high school and continues in college, thus leading to lower male enrollments and test scores over time (Conger and Long 2010). Goldin, Katz, and Kuziemko analyzed a similar issue a few years prior, in which they found that females begin preparing for college earlier than males through test scores and other means (Goldin, Katz, and Kuziemko 2006).

Furthermore, they found that females tend to see college as an opportunity to earn more money that has not been available to women in previous years, thus leading to greater collegiate enrollments among females (Goldin, Katz, and Kuziemko 2006). Though these studies appear to suggest that the opportunities among females in education today are skyrocketing, Jerry Jacobs is quick to point out in his article "Gender Equality and Education" that females are not enrolling in engineering or more high ranking schools as often as males are, nor are they enrolling in doctoral programs as often as males (Jacobs 1996). In his analysis, Jacobs provides a summary of different research articles to provide a more complete understanding of how the educational system works regarding gender (Jacobs 1996). In 2014, Lopez and Barrera wrote a summary article for the Pew Research Center detailing the results of a study regarding the issue of male and female enrollment and race (Lopez and Barrera 2014). Lopez and Barrera mention that disaggregating the male and female enrollments by race will show that differing races have differently sized enrollment gaps between males and females (Lopez and Barrera 2014). Among white enrollments, black enrollments, Asian enrollments, and Hispanic enrollments, Lopez and Barrera explain that Hispanic and black enrollments have the largest male and female enrollment gaps over time (Lopez and Barrera 2014). In terms of disaggregating the data, however, none of these articles have looked at the male and female enrollment gap over time, disaggregating the ratios for state comparison.

In this paper, the rate at which the ten largest states by population are continuously attempting to close the educational gap between male and female enrollment rates over time will be analyzed over a thirty-three-year period, and patterns between the enrollment ratios will be given graphically. In part,

this is to see whether or not the aggregated dataset is representative of the individual states, but it is also to see if any of the states contrast significantly enough from the others to warrant a closer inspection of its geographic factors, policies, or other attributes. The states that are examined here are California, Texas, Florida, New York, Pennsylvania, Illinois, Ohio, Georgia, North Carolina, and Michigan from 1984 to 2016.

### **Data**

The data comes from the *Digest of Education Statistics* reported by the National Center for Education Statistics. This digest is published annually and contains educational data and educational data tables. For the purpose of this analysis, only information on full-time bachelor's students are collected from this data source. This should help to avoid double counting for students who wish to undertake graduate school, since more males or females enrolling in graduate school could bias the data. Furthermore, only the Fall semesters are counted. With this incremental time series data, it is a simple matter to compare and contrast the differing processes. The 2004 issue of the *Digest of Education Statistics* was never produced, however. The implications of this are that the values for the year 2002 have not been reported, and so a proxy must be calculated for the enrollment data within that year. To estimate the data for 2002, the data from 2001 and 2003 are averaged together for each state, thus giving an estimate for the value of the missing data. The independent variable here is time, while the dependent variable is the enrollment ratio, which is calculated by taking the number of females enrolled divided by the number of males enrolled in each state.

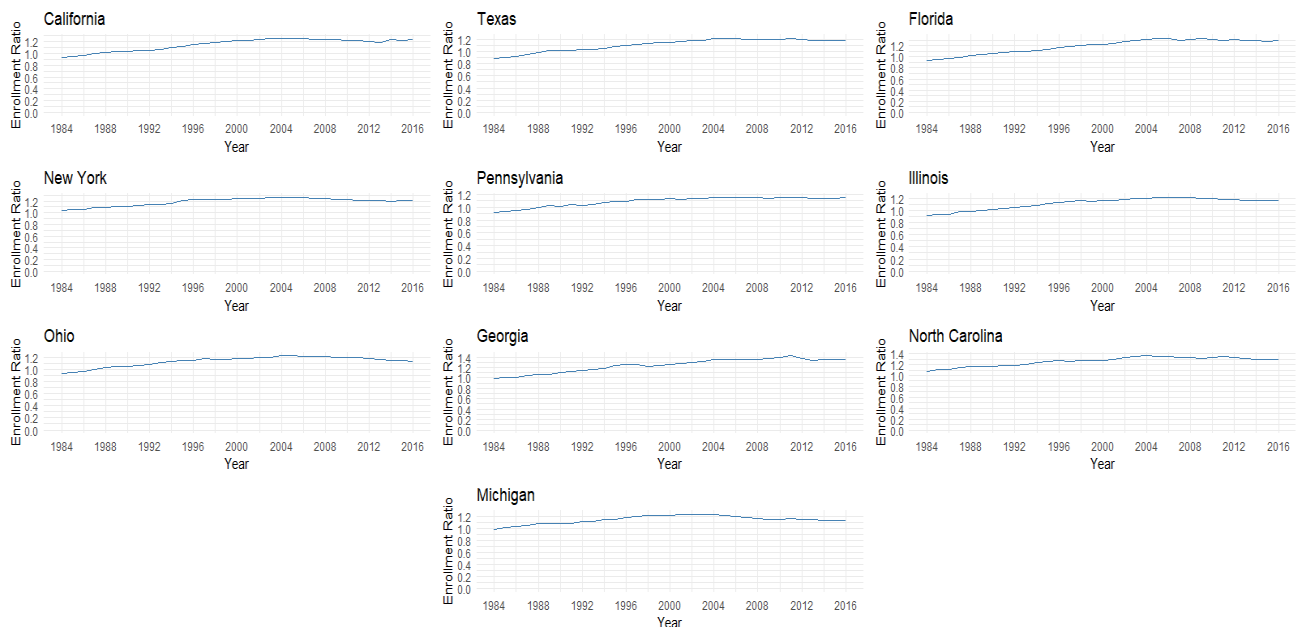
### **Methods**

This analysis begins by examining the graphs of the enrollment ratios for each of the ten states over the thirty – three – year time period. From this point, the data will be checked for stationarity using Augmented Dickey – Fuller tests. If the data series are nonstationary, the graphs of the differenced variables will be provided. If the graphs are stationary, then the next visualization and analysis will be

shown. The autocorrelation and partial autocorrelation functions can then be given in order to better show the dynamics of the variables so that they can be comparatively analyzed. Since there are only thirty-three data points, arima models will not be fit to the data for this analysis. The next visualization consists of a color-coded correlation matrix called a corrpplot. The corrpplot will show the correlations between each of the variables, and thus should be able to show which states are the most different from each other at a single glance. The data can then be visualized, with trend and detrended, by the multivariate time series plot developed by Roger Peng in 2008. Finally, four maps at four points in time (1984, 1994, 2005, and 2016) will be given using choropleth mapping in order to show the possibility that states respond to regional or other types of localized effects. As a conclusion, this analysis will look at the two most different states and suggest reasoning as to why these states are so different. Further methods for future analysis will then be discussed.

## Analysis

### Initial Enrollment Ratio Graphs



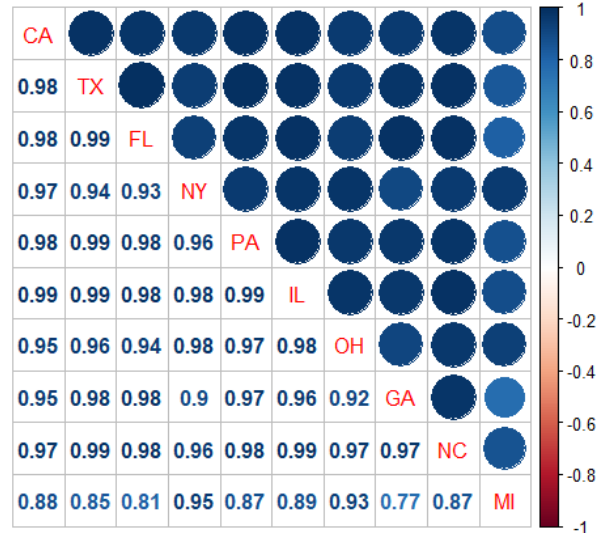
The graphs given above show each states' college enrollment ratios over time. Though each of the functions appear to be similar overall, there are some key differences that allow for comparison. It is important to note that the slow increases or decreases in the data could suggest large changes in the nominal values. Even slight incremental changes to the ratio could signify a change of tens or hundreds of thousands of students. Each of these graphs appear to have an overall increase until around the year 2005, in which they appear to slowly decrease. The only possible exception to this is Georgia, which peaks around 2011 after a large increase in the six – year period before it. The most divergent graph of the ten is Michigan, which decreases around 2002, and never truly stops decreasing. It does have a slight kink, however, during the Great Recession in 2008 in which it briefly increases for about two years before declining once again. Due to the increasing nature of these graphs, it can be difficult to tell by sight if any of them are stationary. It does appear that most of these graphs are stationary, however, there are some states, such as Michigan, which may decrease enough to be nonstationary. Since the enrollment ratios will ideally shift around the number 1.00, any of the states could be nonstationary. In order to better understand if the graphs need to be first differenced or not, Augmented Dickey – Fuller tests must be performed. The number of lags chosen for each of these tests are determined by the Akaike Information Criterion, and the table for the output is given in Appendix A.

The Augmented Dickey – Fuller tests show that there is no evidence of nonstationarity given a 95% confidence level. The only variable that does not pass the test is Georgia if it assumed to have a constant or a constant and a trend. Since all the other the variables reject the null of nonstationarity, it appears that none of them need to be first differenced under this test. It seems likely that Georgia will have a drift term or a drift and a trend, so it is assumed here that Georgia is non stationary. As such, the rest of the data is first differenced for visualization purposes so that the state datasets can be compared more easily. Once the data is first differenced, Georgia passes the Dickey – Fuller test if with a p – value of 0.0005949 if the data is assumed to have no drift term or trend, 0.00008209 if it is assumed to have a

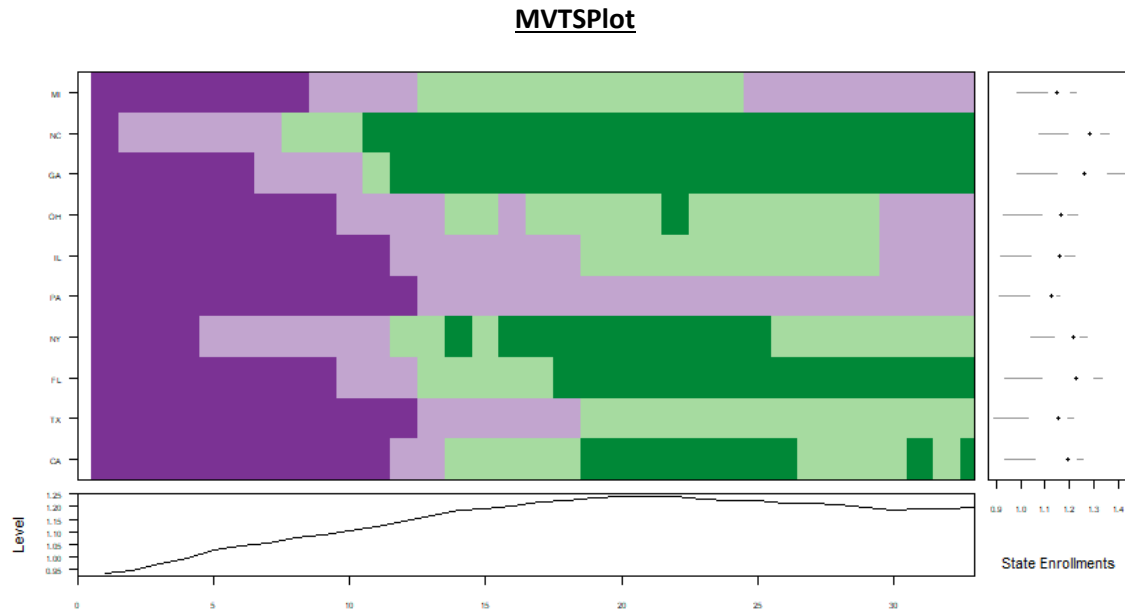
drift term, and 0.00009489 if it is assumed to have both a drift term and a trend. The graphs of the first – differenced enrollment ratios are also given in Appendix A. From these plots, it is shown that Georgia behaves more erratically than the other states, especially around 2011. Around this time, many of the states reached a trough in their enrollment ratio, though some were greater in magnitude than others. In 2011, The enrollment ratio of Georgia is lower than any of the other states at any point in time. At a value of -0.05487, this outlier could be attributable to a reporting error, but, more likely, the value may be due to poor regulation or policy on behalf of the state. In any case, before the conclusion of why this value is so different is addressed, more graphs and charts must be analyzed to see if it could just be a random error.

Now that comparisons have been made and patterns examined for the original graphs, the autocorrelation and partial autocorrelation functions can be examined. In this way, the differences or similarities within the dynamics of each of the states can be analyzed. The correlograms shown in Appendix A appear to have highly persistent autocorrelation functions and a single significant lag in the partial autocorrelation functions, suggesting that each of the states could follow MA(1) processes given more data. For this analysis, arima processes will not be fitted, since the standard errors will be high due to the lack of data. It does appear, however, that each of the variables have similar dynamics based on the autocorrelation and partial autocorrelation functions shown. This may suggest that each of the states follow the same general arima processes over time, though it does not necessarily mean that the coefficients are statistically the same. For this reason, it cannot yet be said that the similar diagnostics of the state level data shows that the aggregate dataset is representative of each of the states. Another way to show this is by using a correlation matrix, in which the correlation between each of the states can be examined to show how closely correlated they are.

### Correlation Plot



The correlation plot given above shows some interesting results. It appears that most of the states are highly correlated, with the exception of Georgia and Michigan. The correlations outside of these two states show mostly strong, positive correlations between 0.99 and 0.97, though there is a correlation of 0.93 between New York and Florida. Including Georgia and Michigan, however, gives a minimum of 0.77. This is not surprising, since the original plot of the data shows that Georgia and Michigan may be two of the most dissimilar states. Michigan differs significantly from the correlation plots shown for each of the other states, with most of the correlations being below a value of 0.9. Georgia is similar in this regard, though it has a minimum value of 0.9 (excluding the correlation with Michigan). Since Michigan has lower values than Georgia, it appears that Michigan is the state that has the largest overall difference in data, with Georgia as the second most different state level dataset. For this analysis, it is assumed that the states with the highest correlation are the states that are most similar. Conversely, the more different two datasets are, the lower the correlation value. In this way, it is possible to quantify the differences in the state datasets using a simple correlation matrix. This will not, however, take any single point in time into consideration. To better understand how well the different periods in time align, a different method, such as a multivariate time series plot, must be considered.



In this style of multivariate time series visualization, originally proposed by Roger Peng in the *Journal of Statistical Software*, each of the states are shown on each row, and each unit of time is measured by each column (Peng, 2008). In doing so, the columns are color coded to show the higher and lower values, in which dark purple represents the first quartile of the dataset, light purple represents the second quartile of the dataset, light green represents the third quartile of the dataset, and dark green represents the fourth quartile of the dataset. In doing this, the multivariate time series plot, as it is called, approximates how closely each state will follow the same general patterns of peaks and troughs as the other nine states by plotting the quartiles of the data. In this way, it functions almost as an intertemporal heat map. The similarities in the datasets can be seen by the way in which the colors cluster at each point in time. The median of each state variable is given on the right as a point above a scale, along with a boxplot representation to approximate how the data is distributed. Furthermore, the median of the variables at each point in time are shown underneath. Thus, the plot shows the similarity in the distributions of the data. It is important to note that the “Level”, as it is called, which is created by taking the median of each time series point, is not scaled from zero. For this reason, it can be misleading to assume that the median of each point is nonstationary. Nonetheless, the first differenced and twice



differenced plots are provided in Appendix B and discussed here so that the possibility of non-stationarity of the plots can be examined.

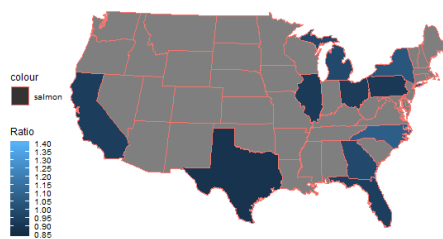
For this first plot, shown above, a significant amount of clustering is seen within the data. Each state plotted shows values within the first and second quartile of the entire set. Pennsylvania has no values in the data above the second quartile, while Michigan, Illinois, and Texas have no values above the third quartile. This may suggest a difference in intercepts, though it could also suggest that these four states behave less erratically and are subject to less fluctuation than the other six. The boxplots on the side of the chart show large differences within the medians, maximums, and minimums. The medians range from 1.237 (Pennsylvania) to 1.281 (North Carolina). The general shape of the boxplots, however, remains relatively the same. Each of the boxplots shows that the data is spaced over larger intervals until the median is reached. After this, the spacing of the data becomes more concentrated, which may suggest a slight upward curve in most of the data. This is confirmed by the "Levels" plot underneath, in which a clear upward movement in the data can be observed. The issue with the assumption that this would suggest a trend begins with the fact that the "Levels" graph is scaled between 0.95 and 1.25, so it is not large enough to conclude that there is a clear trend within the data. Nonetheless, most of the graphs follow this general pattern, with the possible exception of Georgia, which has the widest boxplot and the highest maximum value. The fact that most of the boxplots have different minimums, medians, and maximums appear to suggest different processes occurring within the data. Since Georgia has been shown earlier to be nonstationary, a first differenced version of the multivariate time series plot is given in Appendix B, along with a twice differenced version to ensure nonstationarity. Though the twice differenced version may be excessive for the purposes of this analysis, it is nonetheless provided for the purpose of completeness.

For the first differenced dataset, significant clustering can be seen in the data, particularly near the beginning and end of the series. Here, the medians of each of the series align closer than the

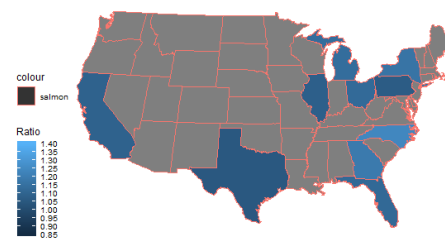
previous plot, between values of 0.004449 (Michigan) and 0.01631 (Georgia). The boxplots still show differences in terms of the maximums and minimums, with Georgia once again covering the widest range. The “Levels” plot underneath is scaled from -0.02 to 0.03, with what appears to be a slight downward pattern in the data. Once again, the scaling may create what looks to be a trend when none exists. To ensure that this possibility is accounted for, however, the twice differenced multivariate time series plot is shown in Appendix B. By the time that the data is differenced twice, there is very little pattern or clustering inherent in the data, though the medians align almost perfectly between -0.0010250 (Illinois) and 0.0033709 (Georgia). The fact that the medians have gradually aligned may be a consequence of differencing. Once again, the minimums and maximums do not align closely, suggesting that there are still some possible distribution differences within the data. Nonetheless, the twice – differenced plot has a higher probability of being over – differenced as compared to the other two. For this reason, the first two plots are given more credibility in this analysis. After the multivariate time series plots have been analyzed, maps can be made at different points in time so that the regional dynamics can be visualized.

### Map Data

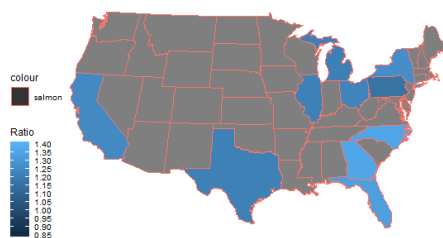
Enrollment Ratio 1984



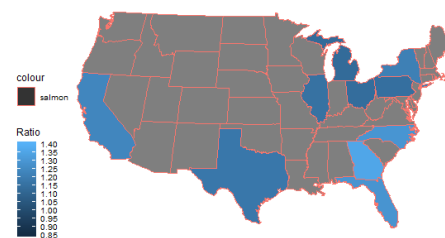
Enrollment Ratio 1994



Enrollment Ratio 2005



Enrollment Ratio 2016



Upon Examining these maps, in which the scale is adjusted so that they can each be read in relation to each other, the general path of the number of enrollments can be observed while the spatial aspects of the data are more visually identifiable than before. Though the geographic, and perhaps, political aspects of the data could be better viewed if the data were on a smaller scale, these four maps should nonetheless provide a fair estimation as to what is occurring in these ten states and at what distance they are from each other. In these maps, each of the states become lighter (signifying a high enrollment ratio) until around 2005, after which they become slightly darker (signifying a lower enrollment ratio). Upon examining Michigan and Georgia, which have been shown throughout this analysis to follow slightly different patterns than the other eight states, it is seen that these two states sit at the center of some regional variation. Michigan sits at the center of some of the Northeastern states, staying nearly the same color as the states surrounding it. Likewise, Georgia sits at the center of some of the southeastern states and appears to influence the states around it to be slightly lighter over time. This can be seen by the fact that the southeastern region stays much lighter than the other regions while Georgia sits at the center of the region. For this reason, in later studies it could be interesting to examine the college enrollment ratios using a gravity model, taking all the other states into account. Nonetheless, it does appear that there is a slight regional effect here that should be noted for future analysis.

### **Conclusion**

After examining the graphs and charts provided in this paper, it is a rational conclusion to suggest that the states' enrollment ratios are subject to statewide and regional fluctuations, suggesting that they may be influenced by geographic, political, economic, or other localized factors. In this analysis, it appears that Georgia and Michigan continuously deviated from the rest of the data more than the other states. Though the purpose of this analysis is to determine if these states follow the same basic patterns under the aggregated dataset, and not to comprehend the reasoning behind differences

within the states, there are suggestions that may be noted here for future analysis. As established in the maps shown previously, there could be regional or cultural factors that Michigan and Georgia are at the center of. These regional or cultural factors may attract different types of students which could, in turn, alter the female/male ratio. Alternatively, the type of colleges and universities in Michigan and Georgia may attract different people on their own. As mentioned previously, Jerry Jacobs discussed in his article that, while females are enrolling in larger numbers than males, the inflow of females is mostly taking place within certain types of schools and majors (Jacobs 1996). Females are enrolling more in liberal arts schools and less in engineering schools, for instance (Jacobs 1996). Therefore, it is possible that the low ratio of females to males seen in Michigan and the high ratios of females to males seen in Georgia stem from the two states having vastly different programs, colleges, and majors. There are also different political leanings and programs between the two states. In 2011, when the first differenced graphs show that Georgia had reached its lowest ratio of  $-0.05487$ , a program had been implemented in the state called the "Complete College Georgia Initiative", in which Georgia used a million dollar grant to reform its educational policies to influence more students to enroll in and stay in college (Governor's Office of Student Achievement 2011). Many of the programs and policies to be restructured did not begin to take effect until 2012, leading to a drastic increase in the number of females in 2012 and 2013 (Governor's Office of Student Achievement 2011). As part of this policy, a scholarship program was created so that lower income middle school students could receive more aid after completing high school coursework (Governor's Office of Student Achievement 2011). Looking at this fact in conjunction with the research article written by Goldin, Katz, and Kuziemko, which stated that females begin preparing sooner for college in high school than males do, it can be noted that females are more likely to take advantage of this opportunity, leading to more long term female enrollment growth (Goldin, Katz, and Kuziemko 2006). Combined with other factors that Georgia implemented with the plan, it seems reasonable that this had an overall positive effect on the number of female enrollments as compared to the number of

male enrollments. It is probable that many other policy factors could have affected the enrollment ratios of these two states, and they may be researched for future analyses, though that is beyond the scope of this current article.

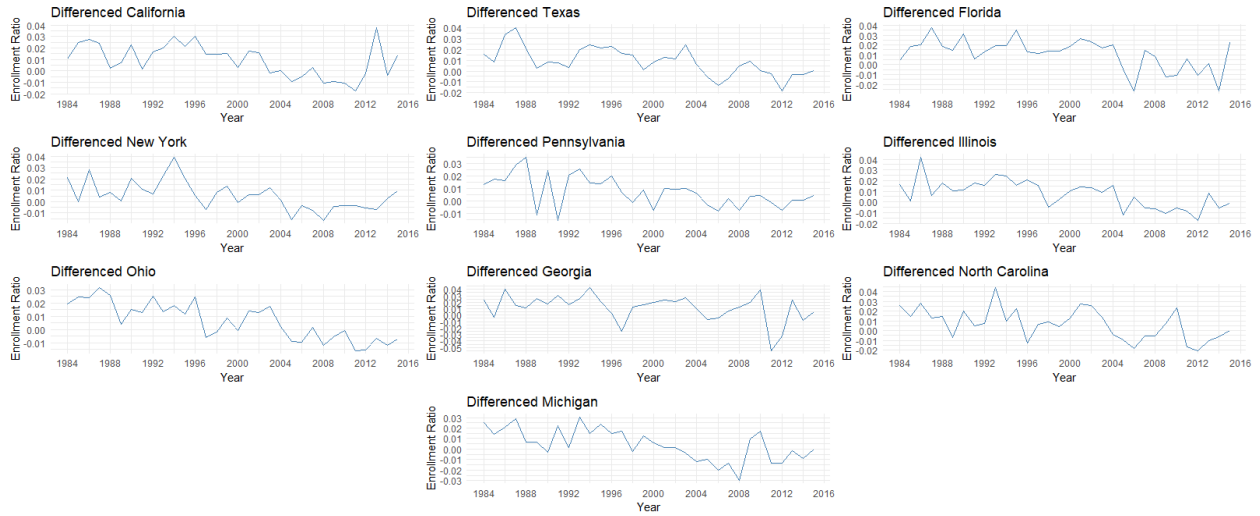
In this study, it is found that the aggregate dataset is generally representative of each of the states, as they all tend to follow the same basic patterns. For some states, however, such as Michigan and Georgia, it is found that the aggregated dataset does not capture the overall patterns that are occurring. It is likely that these states are influenced by regional, geographic, economic, or other localized factors that are not accounted for in the larger dataset. There are, however, limitations to this study. Though graphical analysis and visualization have been given here, no empirical way of directly testing the hypothesis except through multilevel modeling has been found. There are models that could have been used here but have been omitted for the assumptions that they make. A seemingly unrelated regression model could be run and Theil's F – test could be used to test if each dataset has statistically the same coefficients. The issue with this is that the F – test would only show if the effects of the predictor variables are the same, and not the processes themselves. For this reason, the seemingly unrelated regression structure is not used here, nor is it necessarily recommended for future analysis. The best way to empirically test if the states follow the same patterns for future analyses is to set up a multilevel model and use forest plots to determine if the processes can be said to be statistically different. It may also help, in future years, to gather more data to reduce the standard errors of the analysis. After this is performed, differing policy and economic analyses can be performed to understand why the states follow different policies, especially considering that Michigan and Georgia have been shown to follow different processes in this analysis. These would be interesting analyses to run in the future, and they could help to reform or restructure the educational system so that it benefits everyone equally.

## Appendix A (Initial Stationarity Tests and Diagnostics)

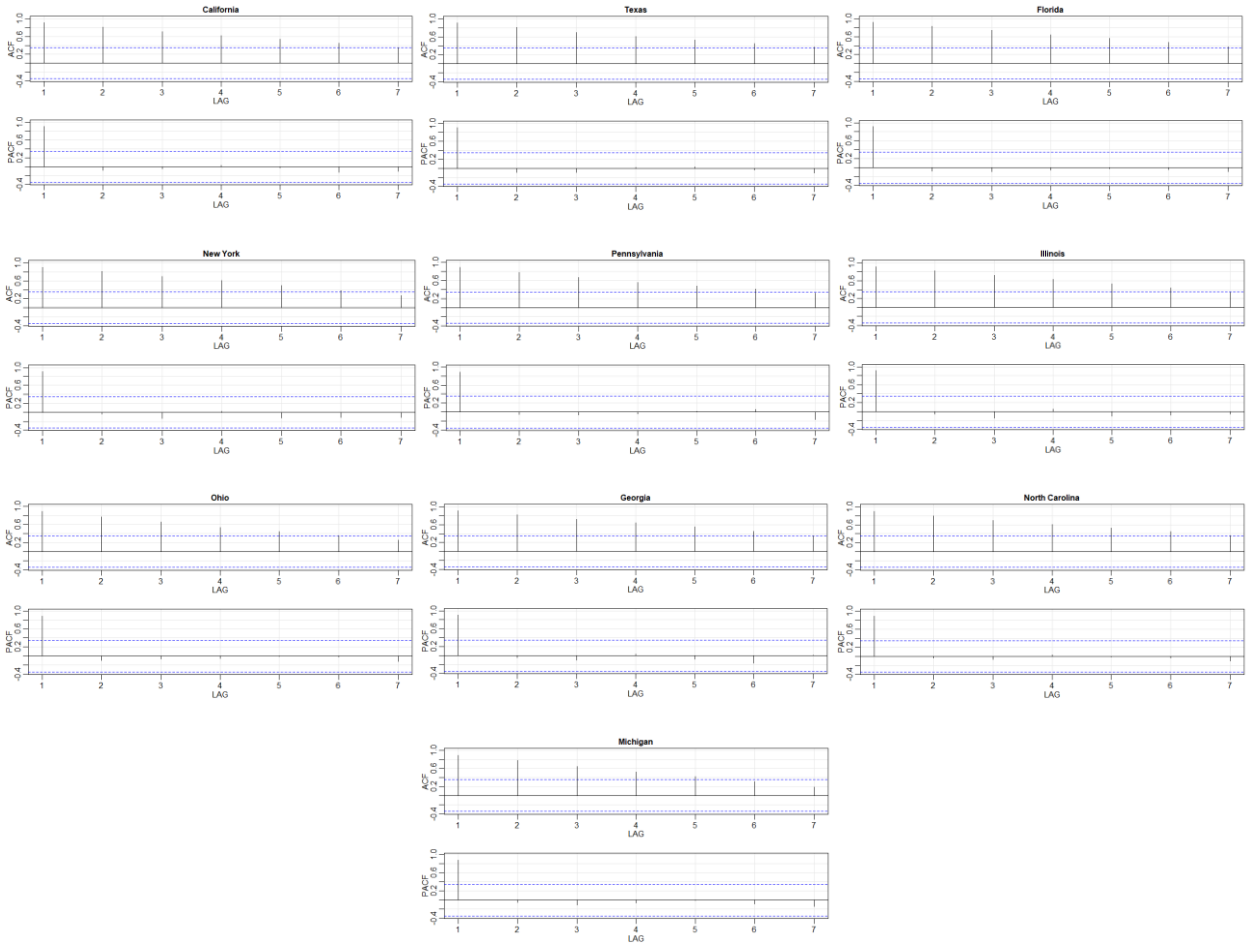
Augmented Dickey – Fuller Test

	OVERALL VALUE OF THE TEST STATISTIC			5 PCT CRITICAL VALUE			SIGNIFICANCE LEVEL	REJECT/FAIL TO REJECT
	Tau	Phi1	Phi2	Tau	Phi1	Phi2	P - Value	Null
CAL (NO TREND OR DRIFT)	1.6339	N/A	N/A	-1.95	N/A	N/A	0.0003612	Reject
CAL (WITH DRIFT)	-2.5808	5.4213	N/A	-2.93	4.86	N/A	0.002575	Reject
CAL (WITH TREND)	-1.3029	3.4873	3.2143	-3.50	5.13	6.73	0.008465	Reject
TEX (NO TREND OR DRIFT)	0.9671	N/A	N/A	-1.95	N/A	N/A	0.000001416	Reject
TEX (WITH DRIFT)	-2.5757	4.1546	N/A	-2.93	4.86	N/A	0.0000179	Reject
TEX (WITH TREND)	-1.0167	2.679	3.2104	-3.50	5.13	6.73	0.0000845	Reject
FLO (NO TREND OR DRIFT)	1.9751	N/A	N/A	-1.95	N/A	N/A	0.0007329	Reject
FLO (WITH DRIFT)	-2.9445	7.7268	N/A	-2.93	4.86	N/A	0.004278	Reject
FLO (WITH TREND)	0.0534	5.6747	5.0759	-3.50	5.13	6.73	0.007176	Reject
NY (NO TREND OR DRIFT)	1.0558	N/A	N/A	-1.95	N/A	N/A	0.005209	Reject
NY (WITH DRIFT)	-2.028	2.8054	N/A	-2.93	4.86	N/A	0.0063	Reject
NY (WITH TREND)	-0.8408	2.033	2.3118	-3.50	5.13	6.73	0.01461	Reject
PENN (NO TREND OR DRIFT)	2.1085	N/A	N/A	-1.95	N/A	N/A	0.02229	Reject
PENN (WITH DRIFT)	-3.9875	12.153	N/A	-2.93	4.86	N/A	0.001549	Reject
PENN (WITH TREND)	-1.3886	8.1307	8.0665	-3.50	5.13	6.73	0.004202	Reject
ILL (NO TREND OR DRIFT)	1.4062	N/A	N/A	-1.95	N/A	N/A	0.002533	Reject
ILL (WITH DRIFT)	-2.8864	5.8262	N/A	-2.93	4.86	N/A	0.002718	Reject
ILL (WITH TREND)	0.2824	6.248	7.4563	-3.50	5.13	6.73	0.000889	Reject
OH (NO TREND OR DRIFT)	0.4805	N/A	N/A	-1.95	N/A	N/A	0.00001131	Reject
OH (WITH DRIFT)	-2.8793	4.3976	N/A	-2.93	4.86	N/A	0.000004567	Reject
OH (WITH TREND)	-0.5345	6.218	9.0138	-3.50	5.13	6.73	0.0000008396	Reject
GEO (NO TREND OR DRIFT)	1.8745	N/A	N/A	-1.95	N/A	N/A	0.02464	Reject
GEO (WITH DRIFT)	-2.0681	4.6521	N/A	-2.93	4.86	N/A	0.09245	Fail to Reject
GEO (WITH TREND)	-0.8528	3.0192	2.0992	-3.50	5.13	6.73	0.1918	Fail to Reject
NC (NO TREND OR DRIFT)	1.1464	N/A	N/A	-1.95	N/A	N/A	0.01647	Reject
NC (WITH DRIFT)	-2.3045	3.6038	N/A	-2.93	4.86	N/A	0.01232	Reject
NC (WITH TREND)	-0.6531	2.423	2.7113	-3.50	5.13	6.73	0.03084	Reject
MIC (NO TREND OR DRIFT)	0.5588	N/A	N/A	-1.95	N/A	N/A	0.009935	Reject
MIC (WITH DRIFT)	-2.1207	2.4969	N/A	-2.93	4.86	N/A	0.003394	Reject
MIC (WITH TREND)	-0.9705	3.0508	4.3043	-3.50	5.13	6.73	0.002124	Reject

### First – Differenced Data

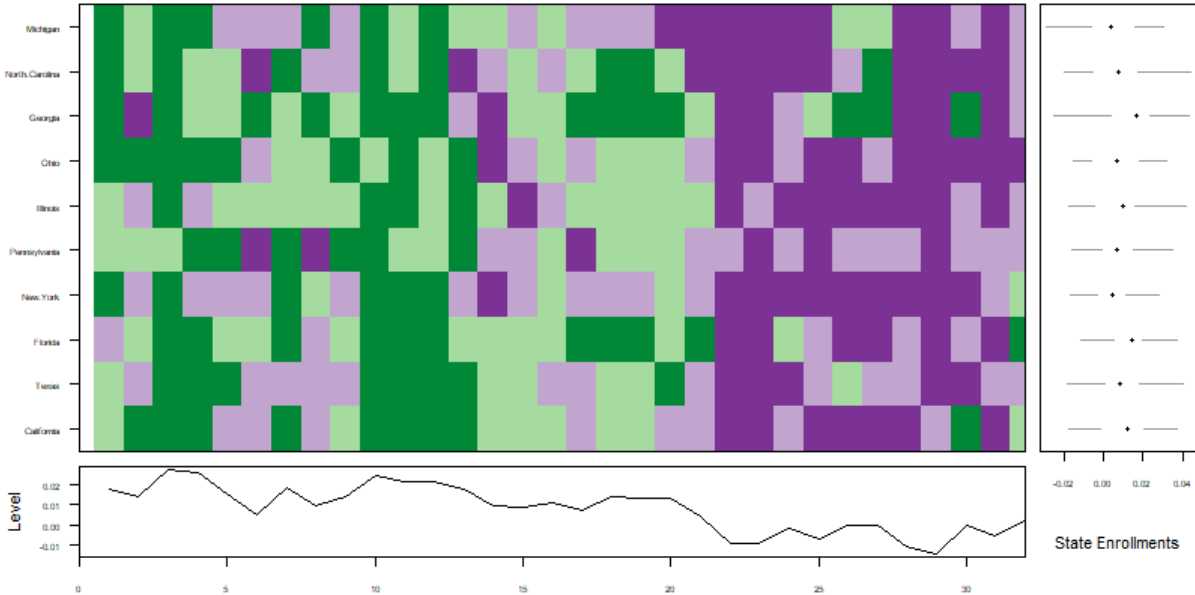


### Autocorrelation and Partial Autocorrelation Functions

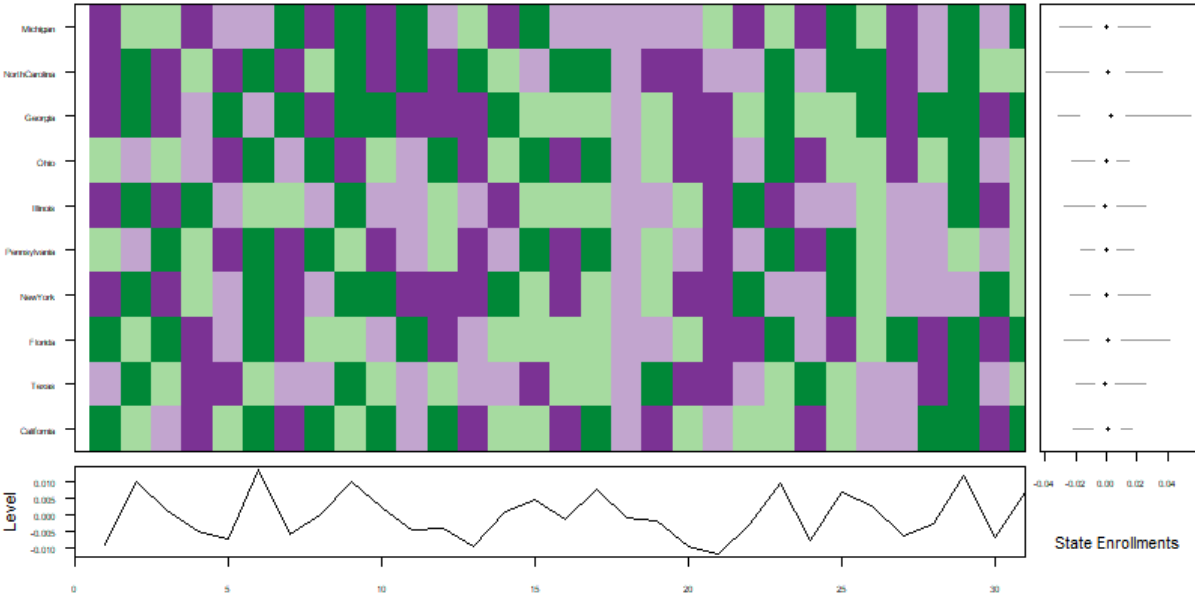


Appendix B (Multivariate Time Series Plots)

First Differenced MVTSPlot



Twice Differenced MVTSPlot





## Appendix C (RStudio Code)

```

getwd()
setwd("C:/Users/Colli/Documents/Data Visualization/Project/")

data1 <- read.csv("Data_Visualization_Input_Data.csv")
data2 <- ts(read.csv("Data_Visualization_Input_Data.csv", header=TRUE),
            start=1984, freq=1)
data3 <- read.csv("Diff_Data_Visualization_Input_Data.csv")
data4 <- ts(read.csv("Diff_Data_Visualization_Input_Data.csv"))
data5 <- read.csv("Data_Visualization_Input_Data_2.csv")
data6 <- read.csv("Diff_Data_Plotting.csv")
data7 <- read.csv("Diff_Diff_Data_Plotting.csv")
data8 <- read.csv("Diff_Diff_Data_Visualization_Input_Data3.csv")

Total <- data1$Total
California <- data1$California
Texas <- data1$Texas
Florida <- data1$Florida
NewYork <- data1$New.York
Pennsylvania <- data1$Pennsylvania
Illinois <- data1$Illinois
Ohio <- data1$Ohio
Georgia <- data1$Georgia
NorthCarolina <- data1$North.Carolina
Michigan <- data1$Michigan
Average <- data1$Average

DCalifornia <- diff(California)
DTexas <- diff(Texas)
DFlorida <- diff(Florida)
DNewYork <- diff(NewYork)
DPennsylvania <- diff(Pennsylvania)
DIllinois <- diff(Illinois)
DOhio <- diff(Ohio)
DGeorgia <- diff(Georgia)
DNorthCarolina <- diff(NorthCarolina)
DMichigan <- diff(Michigan)
DAverage <- diff(Average)

# Initial Analysis
library("astsa")
library("urca")
# Original Data
cal1 <- plot.ts(ts(California, start = 1984, frequency = 1), main = "California", xlab = "Year", ylab = "Enrollment Rate")
cal2 <- acf2(California, main = "California")
df1C <- ur.df(California, type = "none", selectlags = "AIC")
summary(df1C)
df2C <- ur.df(California, type = "drift", selectlags = "AIC")
summary(df2C)
df3C <- ur.df(California, type = "trend", selectlags = "AIC")
summary(df3C)

tex1 <- plot.ts(ts(Texas, start = 1984, frequency = 1), main = "Texas", xlab = "Year", ylab = "Enrollment Rate")
tex2 <- acf2(Texas, main = "Texas")
df1T <- ur.df(Texas, type = "none", selectlags = "AIC")

```

```
summary(df1T)
df2T <- ur.df(Texas, type = "drift", selectlags = "AIC")
summary(df2T)
df3T <- ur.df(Texas, type = "trend", selectlags = "AIC")
summary(df3T)

flo1 <- plot.ts(ts(Florida, start = 1984, frequency = 1), main = "Florida", xlab = "Year", ylab = "Enrollment Rate")
flo2 <- acf2(Florida, main = "Florida")
df1F <- ur.df(Florida, type = "none", selectlags = "AIC")
summary(df1F)
df2F <- ur.df(Florida, type = "drift", selectlags = "AIC")
summary(df2F)
df3F <- ur.df(Florida, type = "trend", selectlags = "AIC")
summary(df3F)

new1 <- plot.ts(ts(NewYork, start = 1984, frequency = 1), main = "New York", xlab = "Year", ylab = "Enrollment Rate")
new2 <- acf2(NewYork, main = "New York")
df1NY <- ur.df(NewYork, type = "none", selectlags = "AIC")
summary(df1NY)
df2NY <- ur.df(NewYork, type = "drift", selectlags = "AIC")
summary(df2NY)
df3NY <- ur.df(NewYork, type = "trend", selectlags = "AIC")
summary(df3NY)

pen1 <- plot.ts(ts(Pennsylvania, start = 1984, frequency = 1), main = "Pennsylvania", xlab = "Year", ylab = "Enrollment Rate")
pen2 <- acf2(Pennsylvania, main = "Pennsylvania")
df1P <- ur.df(Pennsylvania, type = "none", selectlags = "AIC")
summary(df1P)
df2P <- ur.df(Pennsylvania, type = "drift", selectlags = "AIC")
summary(df2P)
df3P <- ur.df(Pennsylvania, type = "trend", selectlags = "AIC")
summary(df3P)

ill1 <- plot.ts(ts(Illinois, start = 1984, frequency = 1), main = "Illinois", xlab = "Year", ylab = "Enrollment Rate")
ill2 <- acf2(Illinois, main = "Illinois")
df1I <- ur.df(Illinois, type = "none", selectlags = "AIC")
summary(df1I)
df2I <- ur.df(Illinois, type = "drift", selectlags = "AIC")
summary(df2I)
df3I <- ur.df(Illinois, type = "trend", selectlags = "AIC")
summary(df3I)

ohi1 <- plot.ts(ts(Ohio, start = 1984, frequency = 1), main = "Ohio", xlab = "Year", ylab = "Enrollment Rate")
ohi2 <- acf2(Ohio, main = "Ohio")
df1O <- ur.df(Ohio, type = "none", selectlags = "AIC")
summary(df1O)
df2O <- ur.df(Ohio, type = "drift", selectlags = "AIC")
summary(df2O)
df3O <- ur.df(Ohio, type = "trend", selectlags = "AIC")
summary(df3O)

geo1 <- plot.ts(ts(Georgia, start = 1984, frequency = 1), main = "Georgia", xlab = "Year", ylab = "Enrollment Rate")
geo2 <- acf2(Georgia, main = "Georgia")
df1G <- ur.df(Georgia, type = "none", selectlags = "AIC")
summary(df1G)
df2G <- ur.df(Georgia, type = "drift", selectlags = "AIC")
summary(df2G)
df3G <- ur.df(Georgia, type = "trend", selectlags = "AIC")
```

```

summary(df3G)

nor1 <- plot.ts(ts(NorthCarolina, start = 1984, frequency = 1), main = "North Carolina", xlab = "Year", ylab = "Enrollment Rate")
nor2 <- acf2(NorthCarolina, main = "North Carolina")
df1NC <- ur.df(NorthCarolina, type = "none", selectlags = "AIC")
summary(df1NC)
df2NC <- ur.df(NorthCarolina, type = "drift", selectlags = "AIC")
summary(df2NC)
df3NC <- ur.df(NorthCarolina, type = "trend", selectlags = "AIC")
summary(df3NC)

mic1 <- plot.ts(ts(Michigan, start = 1984, frequency = 1), main = "Michigan", xlab = "Year", ylab = "Enrollment Rate")
mic2 <- acf2(Michigan, main = "Michigan")
df1M <- ur.df(Michigan, type = "none", selectlags = "AIC")
summary(df1M)
df2M <- ur.df(Michigan, type = "drift", selectlags = "AIC")
summary(df2M)
df3M <- ur.df(Michigan, type = "trend", selectlags = "AIC")
summary(df3M)

# Differenced Data
dcal1 <- plot.ts(ts(DCalifornia, start = 1984, frequency = 1), main = "Differenced California", xlab = "Year", ylab = "Enrollment Rate")
dcal2 <- acf2(DCalifornia, main = "Differenced California")
df1DC <- ur.df(DCalifornia, type = "none", lags = 0)
summary(df1DC)
df2DC <- ur.df(DCalifornia, type = "drift", lags = 0)
summary(df2DC)
df3DC <- ur.df(DCalifornia, type = "trend", lags = 0)
summary(df3DC)

dtex1 <- plot.ts(ts(DTexas, start = 1984, frequency = 1), main = "Differenced Texas", xlab = "Year", ylab = "Enrollment Rate")
dtex2 <- acf2(DTexas, main = "Differenced Texas")
df1DT <- ur.df(DTexas, type = "none", lags = 0)
summary(df1DT)
df2DT <- ur.df(DTexas, type = "drift", lags = 0)
summary(df2DT)
df3DT <- ur.df(DTexas, type = "trend", lags = 0)
summary(df3DT)

dflo1 <- plot.ts(ts(DFlorida, start = 1984, frequency = 1), main = "Differenced Florida", xlab = "Year", ylab = "Enrollment Rate")
dflo2 <- acf2(DFlorida, main = "Differenced Florida")
df1DF <- ur.df(DFlorida, type = "none", lags = 0)
summary(df1DF)
df2DF <- ur.df(DFlorida, type = "drift", lags = 0)
summary(df2DF)
df3DF <- ur.df(DFlorida, type = "trend", lags = 0)
summary(df3DF)

dnew1 <- plot.ts(ts(DNewYork, start = 1984, frequency = 1), main = "Differenced New York", xlab = "Year", ylab = "Enrollment Rate")
dnew2 <- acf2(DNewYork, main = "Differenced New York")
df1DNY <- ur.df(DNewYork, type = "none", lags = 0)
summary(df1DNY)
df2DNY <- ur.df(DNewYork, type = "drift", lags = 0)
summary(df2DNY)
df3DNY <- ur.df(DNewYork, type = "trend", lags = 0)
summary(df3DNY)

```

```
dpen1 <- plot.ts(ts(DPennsylvania, start = 1984, frequency = 1), main = "Differenced Pennsylvania", xlab = "Year", ylab =
"Enrollment Rate")
dpen2 <- acf2(DPennsylvania, main = "Differenced Pennsylvania")
df1DP <- ur.df(DPennsylvania, type = "none", lags = 0)
summary(df1DP)
df2DP <- ur.df(DPennsylvania, type = "drift", lags = 0)
summary(df2DP)
df3DP <- ur.df(DPennsylvania, type = "trend", lags = 0)
summary(df3DP)

dill1 <- plot.ts(ts(DIllinois, start = 1984, frequency = 1), main = "Differenced Illinois", xlab = "Year", ylab = "Enrollment Rate")
dill2 <- acf2(DIllinois, main = "Differenced Illinois")
df1DI <- ur.df(DIllinois, type = "none", lags = 0)
summary(df1DI)
df2DI <- ur.df(DIllinois, type = "drift", lags = 0)
summary(df2DI)
df3DI <- ur.df(DIllinois, type = "trend", lags = 0)
summary(df3DI)

dohi1 <- plot.ts(ts(DOhio, start = 1984, frequency = 1), main = "Differenced Ohio", xlab = "Year", ylab = "Enrollment Rate")
dohi2 <- acf2(DOhio, main = "Differenced Ohio")
df1DO <- ur.df(DOhio, type = "none", lags = 0)
summary(df1DO)
df2DO <- ur.df(DOhio, type = "drift", lags = 0)
summary(df2DO)
df3DO <- ur.df(DOhio, type = "trend", lags = 0)
summary(df3DO)

dgeo1 <- plot.ts(ts(DGeorgia, start = 1984, frequency = 1), main = "Differenced Georgia", xlab = "Year", ylab = "Enrollment
Rate")
dgeo2 <- acf2(DGeorgia, main = "Differenced Georgia")
df1DG <- ur.df(DGeorgia, type = "none", lags = 0)
summary(df1DG)
df2DG <- ur.df(DGeorgia, type = "drift", lags = 0)
summary(df2DG)
df3DG <- ur.df(DGeorgia, type = "trend", lags = 0)
summary(df3DG)

dnor1 <- plot.ts(ts(DNorthCarolina, start = 1984, frequency = 1), main = "Differenced North Carolina", xlab = "Year", ylab =
"Enrollment Rate")
dnor2 <- acf2(DNorthCarolina, main = "Differenced North Carolina")
df1DNC <- ur.df(DNorthCarolina, type = "none", lags = 0)
summary(df1DNC)
df2DNC <- ur.df(DNorthCarolina, type = "drift", lags = 0)
summary(df2DNC)
df3DNC <- ur.df(DNorthCarolina, type = "trend", lags = 0)
summary(df3DNC)

dmic1 <- plot.ts(ts(DMichigan, start = 1984, frequency = 1), main = "Differenced Michigan", xlab = "Year", ylab = "Enrollment
Rate")
dmic2 <- acf2(DMichigan, main = "Differenced Michigan")
df1DM <- ur.df(DMichigan, type = "none", lags = 0)
summary(df1DM)
df2DM <- ur.df(DMichigan, type = "drift", lags = 0)
summary(df2DM)
df3DM <- ur.df(DMichigan, type = "trend", lags = 0)
summary(df3DM)
```

```

# 1st Plot of the Data
library("ggplot2")
g1cal <- ggplot(data = data1, mapping = aes(x = Year, y = California, color = Value)) +
  ggtitle("California") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1cal

g1tex <- ggplot(data = data1, mapping = aes(x = Year, y = Texas, color = Value)) +
  ggtitle("Texas") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1tex

g1flo <- ggplot(data = data1, mapping = aes(x = Year, y = Florida, color = Value)) +
  ggtitle("Florida") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1flo

g1new <- ggplot(data = data1, mapping = aes(x = Year, y = NewYork, color = Value)) +
  ggtitle("New York") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1new

g1pen <- ggplot(data = data1, mapping = aes(x = Year, y = Pennsylvania, color = Value)) +
  ggtitle("Pennsylvania") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)

```

g1pen

```
g1ill <- ggplot(data = data1, mapping = aes(x = Year, y = Illinois, color = Value)) +
  ggtitle("Illinois") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1ill
```

```
g1ohi <- ggplot(data = data1, mapping = aes(x = Year, y = Ohio, color = Value)) +
  ggtitle("Ohio") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1ohi
```

```
g1geo <- ggplot(data = data1, mapping = aes(x = Year, y = Georgia, color = Value)) +
  ggtitle("Georgia") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1geo
```

```
g1nor <- ggplot(data = data1, mapping = aes(x = Year, y = NorthCarolina, color = Value)) +
  ggtitle("North Carolina") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1nor
```

```
g1mic <- ggplot(data = data1, mapping = aes(x = Year, y = Michigan, color = Value)) +
  ggtitle("Michigan") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(0), max(1.45), by = 0.2),1)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1mic
```

```
library("gridExtra")
```

```
lay1 <- rbind(c(1, 2, 3),
             c(4, 5, 6),
             c(7, 8, 9),
             c(NA, 10, NA))
grid.arrange(g1cal, g1tex, g1flo, g1new, g1pen, g1ill, g1ohi, g1geo, g1nor, g1mic, layout_matrix = lay1)
```

```
acf2(California, main = "California")
acf2(Texas, main = "Texas")
acf2(Florida, main = "Florida")
acf2(NewYork, main = "New York")
acf2(Pennsylvania, main = "Pennsylvania")
acf2(Illinois, main = "Illinois")
acf2(Ohio, main = "Ohio")
acf2(Georgia, main = "Georgia")
acf2(NorthCarolina, main = "North Carolina")
acf2(Michigan, main = "Michigan")
```

```
# Differenced Plots
```

```
library("ggplot2")
g1cald <- ggplot(data = data6, mapping = aes(x = Year, y = DiffCalifornia, color = Value)) +
  ggtitle("Differenced California") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1cald
```

```
g1texd <- ggplot(data = data6, mapping = aes(x = Year, y = DiffTexas, color = Value)) +
  ggtitle("Differenced Texas") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1texd
```

```
g1flod <- ggplot(data = data6, mapping = aes(x = Year, y = DiffFlorida, color = Value)) +
  ggtitle("Differenced Florida") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1flod
```

```
g1newd <- ggplot(data = data6, mapping = aes(x = Year, y = DiffNewYork, color = Value)) +
  ggtitle("Differenced New York") +
  xlab("Year") +
```

```

ylab("Enrollment Ratio") +
geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
theme_minimal() +
scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
expand_limits(x=1984, y=0)
g1newd

g1pend <- ggplot(data = data6, mapping = aes(x = Year, y = DiffPennsylvania, color = Value)) +
  ggtitle("Differenced Pennsylvania") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1pend

g1illd <- ggplot(data = data6, mapping = aes(x = Year, y = DiffIllinois, color = Value)) +
  ggtitle("Differenced Illinois") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1illd

g1ohid <- ggplot(data = data6, mapping = aes(x = Year, y = DiffOhio, color = Value)) +
  ggtitle("Differenced Ohio") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1ohid

g1geod <- ggplot(data = data6, mapping = aes(x = Year, y = DiffGeorgia, color = Value)) +
  ggtitle("Differenced Georgia") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.55), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1geod

g1nord <- ggplot(data = data6, mapping = aes(x = Year, y = DiffNorthCarolina, color = Value)) +
  ggtitle("Differenced North Carolina") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +

```



```

scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
expand_limits(x=1984, y=0)
g1nord

```

```

g1micd <- ggplot(data = data6, mapping = aes(x = Year, y = DiffMichigan, color = Value)) +
  ggtitle("Differenced Michigan") +
  xlab("Year") +
  ylab("Enrollment Ratio") +
  geom_line(stat="identity", color = "steelblue", show.legend = TRUE) +
  theme_minimal() +
  scale_y_continuous(breaks = round(seq(min(-0.03), max(0.04), by = 0.01),2)) +
  scale_x_continuous(breaks = round(seq(min(1984), max(2016), by = 4),1)) +
  expand_limits(x=1984, y=0)
g1micd

```

```

library("gridExtra")
grid.arrange(g1cald, g1texd, g1flod, g1newd, g1pend, g1illd, g1ohid, g1geod, g1nord, g1micd, layout_matrix = lay1)

```

```

# Correlation Plot
library("corrplot")
corrplot.mixed(cor(data5))

```

```

# Multivariate Time Series Plots
library("mvtspplot")
datmat1 <- data.matrix(data5)
mvtspplot(datmat1, group = NULL, xtime = NULL, norm = "global", levels = 4,
  smooth.df = NULL, margin = TRUE, sort = NULL, main = "State Enrollments",
  palette = "PRGn", rowstat = "median", xlim, bottom.ylim = NULL,
  right.xlim = NULL, gcol = 1)

```

```

summary(Average)
summary(California)
summary(Texas)
summary(Florida)
summary(NewYork)
summary(Pennsylvania)
summary(Illinois)
summary(Ohio)
summary(Georgia)
summary(NorthCarolina)
summary(Michigan)

```

```

datmat2 <- data.matrix(data3)
mvtspplot(datmat2, group = NULL, xtime = NULL, norm = "global", levels = 4,
  smooth.df = NULL, margin = TRUE, sort = NULL, main = "State Enrollments",
  palette = "PRGn", rowstat = "median", xlim, bottom.ylim = NULL,
  right.xlim = NULL, gcol = 1)

```

```

summary(DAverage)
summary(DCalifornia)
summary(DTexas)
summary(DFlorida)

```

```

summary(DNewYork)
summary(DPennsylvania)
summary(DIllinois)
summary(DOhio)
summary(DGeorgia)
summary(DNorthCarolina)
summary(DMichigan)

datmat3 <- data.matrix(data8)
mvtplot(datmat3, group = NULL, xtime = NULL, norm = "global", levels = 4,
        smooth.df = NULL, margin = TRUE, sort = NULL, main = "State Enrollments",
        palette = "PRGn", rowstat = "median", xlim, bottom.ylim = NULL,
        right.xlim = NULL, gcol = 1)

summary(DDAverage)
summary(DDCalifornia)
summary(DDTexas)
summary(DDFlorida)
summary(DDNewYork)
summary(DDPennsylvania)
summary(DDIllinois)
summary(DDOhio)
summary(DDGeorgia)
summary(DDNorthCarolina)
summary(DDMichigan)

# Map
mapdata1 <- read.csv("2MapData2016.csv")
mapdata2 <- read.csv("MapData1984.csv")
mapdata3 <- read.csv("MapData1994.csv")
mapdata4 <- read.csv("MapData2005.csv")

library("maps")
library("dplyr")
library("ggplot2")
library("gridExtra")

dfstate <- map_data("state")
ggplot(dfstate) + geom_polygon(aes(long, lat, group = group), color = "white")
aggdat1 <- left_join(dfstate, mapdata1, by = c("region" = "state"))

map2016 <- ggplot(data = aggdat1) +
  ggtitle("Enrollment Ratio 2016") +
  labs(fill = "Enrollment Ratio") +
  geom_polygon(aes(long, lat, group = group, fill = enrollrat, color = "salmon")) +
  coord_map("bonne", parameters = 10) +
  ggthemes::theme_map() +
  scale_fill_continuous(limits = c(0.85,1.40), breaks = c(0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.40),
  guide = guide_colourbar(title = "Ratio", draw.ulim = FALSE, draw.llim = FALSE))
map2016

aggdat2 <- left_join(dfstate, mapdata2, by = c("region" = "state"))

map1984 <- ggplot(data = aggdat2) +

```

```

ggtitle("Enrollment Ratio 1984") +
labs(fill = "Enrollment Ratio") +
geom_polygon(aes(long, lat, group = group, fill = enrollrat, color = "salmon")) +
coord_map("bonne", parameters = 10) +
ggthemes::theme_map() +
scale_fill_continuous(limits = c(0.85,1.40), breaks = c(0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.40),
guide = guide_colourbar(title = "Ratio", draw.ulim = FALSE, draw.llim = FALSE))
map1984

```

```
aggdat3 <- left_join(dfstate, mapdata3, by = c("region" = "state"))
```

```

map1994 <- ggplot(data = aggdat3) +
ggtitle("Enrollment Ratio 1994") +
labs(fill = "Enrollment Ratio") +
geom_polygon(aes(long, lat, group = group, fill = enrollrat, color = "salmon")) +
coord_map("bonne", parameters = 10) +
ggthemes::theme_map() +
scale_fill_continuous(limits = c(0.85,1.40), breaks = c(0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.40),
guide = guide_colourbar(title = "Ratio", draw.ulim = FALSE, draw.llim = FALSE))
map1994

```

```
aggdat4 <- left_join(dfstate, mapdata4, by = c("region" = "state"))
```

```

map2005 <- ggplot(data = aggdat4) +
ggtitle("Enrollment Ratio 2005") +
labs(fill = "Enrollment Ratio") +
geom_polygon(aes(long, lat, group = group, fill = enrollrat, color = "salmon")) +
coord_map("bonne", parameters = 10) +
ggthemes::theme_map() +
scale_fill_continuous(limits = c(0.85,1.40), breaks = c(0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.40),
guide = guide_colourbar(title = "Ratio", draw.ulim = FALSE, draw.llim = FALSE))
map2005

```

```
grid.arrange(map1984, map1994, map2005, map2016)
```

## Appendix D

### Bibliography

- Agger, Charlotte, Judith Meece, and Soo-yong Byun. 2018. "The Influences of Family and Place on Rural Adolescents' Educational Aspirations and Post-secondary Enrollment." *Journal of Youth and Adolescence* 47 (December 2018): 2554 – 2568.
- Conger, Dylan and Mark C. Long. 2010. "Why Are Men Falling Behind? Gender Gaps in College Performance and Persistence." *The ANNALS of the American Academy of Political and Social Science* 627, no. 1 (January 4): 184 – 214.
- National Center for Education Statistics. (1990-2017). "Total Fall Enrollment in Degree-Granting Postsecondary Institutions, by Attendance Status, Sex, and State or Jurisdiction". *Digest of Education Statistics*. Accessed at: <https://nces.ed.gov/pubsearch/getpubcats.asp?sid=091#061>
- National Center for Education Statistics. (1987-1989). "Total Fall Enrollment in Degree-Granting Postsecondary Institutions, by Attendance Status, Sex, and State or Jurisdiction". *Digest of Education Statistics*. Accessed at: <https://catalog.hathitrust.org/api/volumes/oclc/3133477.html>
- Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20 no. 4 (Fall): 133-156.
- Jacobs, Jerry A. 1996. "Gender Inequality and Higher Education." *Annual Review of Sociology* 22: 153-185.
- Lopez, Mark Hugo and Ana Gonzalez-Barrera. 2014. "Women's College Enrollment Gains Leave Men Behind." *Pew Research Center*, March 6. <http://www.pewresearch.org/fact-tank/2014/03/06/womens-college-enrollment-gains-leave-men-behind/> (February 19, 2019).
- Peng, Roger. 2008. "A Method for Visualizing Multivariate Time Series Data." *Journal of Statistical Software* 25 (March 31): 1 – 17.
- The Governor's Office of Student Achievement. 2011. "Complete College Georgia: An Overview." <https://gosa.georgia.gov/complete-college-georgia-overview> (April 30, 2019).