

December 4, 2018
Senior Seminar: Applied Statistics Research
John Zhang

Factors Influencing Income Per Capita

Introduction

There is a wide variety of theories on what factors truly affect income, with commonly believed factors such as gender, race and educational attainment. However, there have been many changes in society such as the recent proliferation in higher educational attainment that may have changed the significance of each factor. This project aims to analyze the relationship between different societal factors and income per capita on a statewide level from 2015 to 2017 and compare how the strength of these factors change during this time period. I hope to identify the main factors that affect individual income per capita, any trends that occur during the time period of my analysis, and understand why these trends or patterns, if any, occur.

Literature Review

There has been a sizable amount of research that has examined different factors affecting income, and through analyzing previous research I have been able to gain insight on which factors to focus on in my own study.

Geography. The relationship between geographical elements, such as climate zone and distance from the coast, and economic variables such as income per capita and growth has been explored in several different studies by Jeffery Sachs and the National Bureau of Economic Research (Sachs 1). Surprisingly, their research has indicated that although there is strong evidence that geography has an impact in economic growth, geography has little direct effect on income per capita (Sachs 2). However, the study notes that their results on economic development may be inaccurate, as it is very difficult to measure economic development with only a few variables (Sachs 4).

Race. The effect that race demographics has on income is well-known and has received increased scrutiny during these more contentious times, specifically the income gap between whites and blacks. According to a report by the Pew Research Center that analyzed data from the 2014 Census, the median adjusted income for black and white households was \$43,300 and \$71,300 respectively (Pew Research 8). Even when attempting to control for educational attainment, households headed by a black college degree holder on average still only earned \$82,300 while a household headed by a white college degree holder earned \$106,600 (Pew Research 8). These statistics clearly demonstrate a substantial difference between both groups, and I want to explore how racial demographics will affect income per capita in my own study.

Education. Education is another potential factor that I wanted to explore, as it's very intuitive that better education would correlate with higher income per capita. A 1992 study by David Card and Alan Krueger examined how the quality of a school, which is measured by statistics such as pupil/teacher ratio, average term length, and relative teacher pay, can affect the "rate of return" on a student's education, which refers to a student's earnings after schooling (Card and Krueger 1). The study utilized the 1980 census to gather earnings data on a statewide level for men and information from the U.S Office of Education for characteristics of school quality. Ultimately, the study found that men in higher-quality schools would have higher earnings after controlling for other potential aspects related to a student's background (Card and Krueger 36). In particular, results indicated that decreasing the pupil/teacher ratio would raise average future earnings by 4.2%, and that a 30% increase in teacher pay increased average future earnings by 1.34%. However, the authors of the study emphasized that even if an increase in school quality could raise average rate of returns, it would be possible for the distribution of

earnings to remain unchanged, and thus increase wage inequality. It is also important to note that the study only analyzed data on men.

Gender. In a 2007 study by Donna Bobbitt-Zeher, the relationship between gender and income is explored in the context of increasing educational attainment by women. Although it is well documented that women have actually begun to outperform men in educational attainment, women may still not receive the same returns on education that men do (Bobbitt-Zeher 2). The study sought to examine this by utilizing data from NELS, which was a report that followed about 25,000 students from when they started eighth grade through college and recorded their levels of educational attainment and income before taxes through surveys (Bobbitt-Zeher 7). Two main inferences resulted from the study, first being that the choice of college major had an important effect on income, and that majors associated with women seemed to have lower income (Bobbitt-Zeher 15). Second, even if the choice of college major was the same, women still made about \$4,400 less annually than men (Bobbitt-Zeher 15).

Internet. A Pew Research Study by Jacob Poushter in 2016 analyzed technology use in different countries, and how they related to each country's economic characteristics such as GDP and income per capita. Unsurprisingly, there was a strong correlation found between a country's income per capita and internet access, and that internet usage is rapidly growing in developing countries (Poushter 7). There was also a strong correlation, of 0.87, found between a country's GDP per capita adjusted for purchasing power and internet access (Poushter 8). Interestingly, the study also reported that there is a GDP per capita cutoff point where any additional GDP per capita does not affect internet access rates.

Mean Commuting Distance. In a 2016 study by the Institute for Urban Economic Research, researchers examined how commuting distances in Denmark, the UK and the USA

relate to income. The data was taken from national surveys for each country, with the UK data spanning 2002-2012, the US data including 1995, 2001, 2009 and the Denmark data spanning 2001-2010 (Carra, Mulalic, Fosgerau, Barthelemy 2). The study used a power law function to model the relationship between commuting distances and income, and found that in Denmark and the UK, workers with higher incomes had longer commuter distances. The data also demonstrated that the relationship between average commuting distance and income had a slowly decaying tail for longer distances (Carra, Mulalic, Fosgerau, Barthelemy 2). Thus, when comparing datapoints with long distances, their respective incomes did not vary as much.

Each of these studies analyzed the effects that different factors had on both income per capita and household income levels. Out of all these factors that were examined, only geographical elements such as climate zone were found to have no significant effect on income per capita, however these elements were found to have an impact on economic development. Thus, I will not analyze geographical elements in my own study, as I am more focused on factors that affect income per capita than economic growth.

Although I was already generally aware of the effects that demographic factors such as gender and race have on income per capita, I found each of these factor's respective studies to be extremely insightful. In terms of race, since it is well known that different racial groups have varying levels of educational attainment, I was not sure whether race would have a significant effect when holding other variables constant, or if the differing levels of educational attainment between racial groups were the underlying cause of variations in income per capita for each group. The Pew Research study on the relationship between race and income explored that in a section of the study by comparing subjects from different racial groups who all possessed

bachelor's degrees and found that there was still a substantial difference in income. This gave me further confidence in using race as a factor in my analysis.

The study on gender offered similar insight, as when exploring the income gap between genders the study examined how the choice of college majors associated with women generally led to lower income levels, and that even women with the same major or concentration had lower earnings than men.

Although I knew from the start of this study that I wanted to explore the effects that education has on income per capita, I was not sure how to measure educational quality. Thus, the 1992 study by David Card and Alan Krueger was extremely useful in helping me discern how to approach this variable. The study determined that decreases in the pupil/teacher ratio and increases in teacher wages have positive correlations with future income per capita, and these are variables I want to explore in my study.

I found the study on the relationship between internet access to income and GDP to be more straightforward compared to the other studies, as a very clear correlation between internet access and income was found. However, the fact that internet access rates did not increase at the same rate once countries reached a certain GDP per capita level was unexpected and could prove to be relevant in my own study.

The study on mean commuting distances was extremely insightful, as not only was commuting distance a potential variable I was unaware of, but also the study found a positive correlation between mean commuting distance and income in both Denmark and the UK, but not in the US. At first, I found the positive correlation to be unintuitive, as I believed workers with higher incomes could afford housing with a shorter commute, but the study implies that this correlation exists because workers are only willing to commute longer distances if their income

is higher. Although there was no correlation found in the US, I will still analyze if worker commuting has a relationship with income.

Methodology and Data:

The analysis utilized data from several sources, with the majority of the data coming from the US Census.

Type of Data	Title	Measurement	Source	Year
Independent	Black	% of State Population that is Black	United States Census Bureau	2015-2017
Independent	Female	% of State Population who are female	United States Census Bureau	2015-2017
Independent	Pupil/Teacher Ratio	Students Enrolled Per Teacher	National Education Association	2015-2017
Independent	Teacher Wages	Average Salaries of Teachers	National Education Association	2015-2017
Independent	Mean Travel Time	Mean Commuting Time	United States Census Bureau	2015-2017
Independent	Internet Access	Households with a Broadband Internet Connection	United States Census Bureau	2015-2017
Independent	Civilian Labor Force	% of Population in Labor Force	United States Census Bureau	2015-2017
Independent	Now Married	% of Population that is Married	United States Census Bureau	2015-2017
Independent	High School Graduate	% of Population that Graduated from High School	United States Census Bureau	2015-2017
Independent	Some College/Associate's Degree	% of Population with Some College or Associate's Degree	United States Census Bureau	2015-2017
Independent	Bachelor's Degree	% of Population with Bachelor's Degree	United States Census Bureau	2015-2017
Independent	Graduate Degree	% of Population with Graduate Degree	United States Census Bureau	2015-2017
Independent	Foreign Born	Foreign-Born Population	United States Census Bureau	2015-2017
Independent	Family Households	% of Households that are Family Households	United States Census Bureau	2015-2017
Independent	Median Age	Population Median Age	United States Census Bureau	2015-2017

Independent	Aged 18 to 64	% of Population Aged 18 to 64	United States Census Bureau	2015-2017
Dependent	Income Per Capita	Individual Per Capita Income	United States Census Bureau	2015-2017

Due to the large number of variables I was working with, I utilized a function in R (the statistics software I was using) called Regsubsets which determines which subsets of variables form models with the highest adjusted R^2 value for a linear regression.

Interestingly, when I used this function for each year (2015, 2016, 2017), the function outputted the same subsets of variables for each year. I then ran a linear model for each year.

2015 Model: *Per_capita_income ~ Civilian_labor_force + Mean_travel_time + Internet_Access + Teacher_Wages + Pupil_Teacher_Ratio + Black*

<i>Independent Variable</i>	<i>Coefficient Estimate</i>	<i>Standard Error</i>	<i>T-Value</i>	<i>P-Value</i>
<i>Civilian Labor Force</i>	3.365e+02	1.249e+02	2.695	0.009936 *
<i>Mean Travel Time</i>	8.646e+01	1.419e+02	0.609	0.545386
<i>Internet Access</i>	4.036e+02	1.239e+02	3.257	0.002174 *
<i>Teacher Wages</i>	2.181e-01	4.583e-02	4.759	2.13e-05 *
<i>Pupil Teacher Ratio</i>	-4.199e+02	1.155e+02	-3.634	0.000726 *
<i>Black</i>	1.480e+04	4.244e+03	3.487	0.001119 *

Residual standard error: 2069 on 44 degrees of freedom
Multiple R-squared: 0.8544, Adjusted R-squared: 0.8346
F-statistic: 43.04 on 6 and 44 DF, p-value: < 2.2e-16

2016 Model: *Per_capita_income ~ Civilian_labor_force + Mean_travel_time + Internet_Access + Teacher_Wages + Pupil_Teacher_Ratio + Black*

<i>Independent Variable</i>	<i>Coefficient Estimate</i>	<i>Standard Error</i>	<i>T-Value</i>	<i>P-Value</i>
<i>Civilian Labor Force</i>	4.227e+02	1.423e+02	2.971	0.00480 *
<i>Mean Travel Time</i>	2.498e+02	1.414e+02	1.766	0.08429
<i>Internet Access</i>	3.588e+02	1.725e+02	2.080	0.04333 *
<i>Teacher Wages</i>	1.974e-01	4.847e-02	4.074	0.00019 *
<i>Pupil Teacher Ratio</i>	-2.896e+02	1.201e+02	-2.412	0.02012 *
<i>Black</i>	9.378e+03	4.375e+03	2.144	0.03762 *

Residual standard error: 2306 on 44 degrees of freedom

Multiple R-squared: 0.8191, Adjusted R-squared: 0.7944

F-statistic: 33.19 on 6 and 44 DF, p-value: 8.75e-15

2017 Model: *Per_capita_income* ~ *Civilian_labor_force* + *Mean_travel_time* + *Internet_Access* + *Teacher_Wages* + *Pupil_Teacher_Ratio* + *Black*

<i>Independent Variable</i>	<i>Coefficient Estimate</i>	<i>Standard Error</i>	<i>T-Value</i>	<i>P-Value</i>
<i>Civilian Labor Force</i>	4.505e+02	1.352e+02	3.333	0.001749 *
<i>Mean Travel Time</i>	2.890e+02	1.482e+02	1.950	0.057528
<i>Internet Access</i>	3.971e+02	1.763e+02	2.253	0.029325 *
<i>Teacher Wages</i>	2.007e-01	4.920e-02	4.079	0.000187 *
<i>Pupil Teacher Ratio</i>	-2.856e+02	1.242e+02	-2.300	0.026238 *
<i>Black</i>	9.163e+03	4.428e+03	2.069	0.044430 *

Residual standard error: 2380 on 44 degrees of freedom

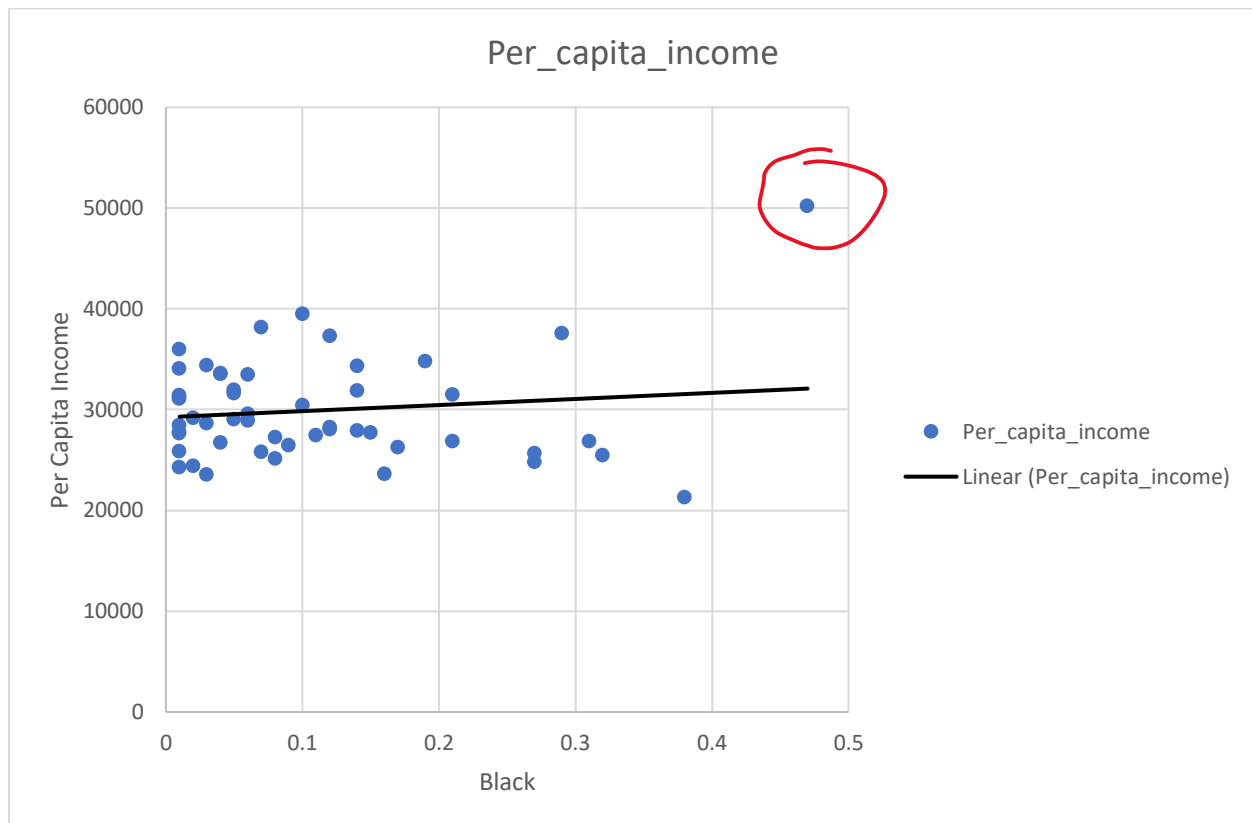
Multiple R-squared: 0.8281, Adjusted R-squared: 0.8047

F-statistic: 35.34 on 6 and 44 DF, p-value: 2.878e-15

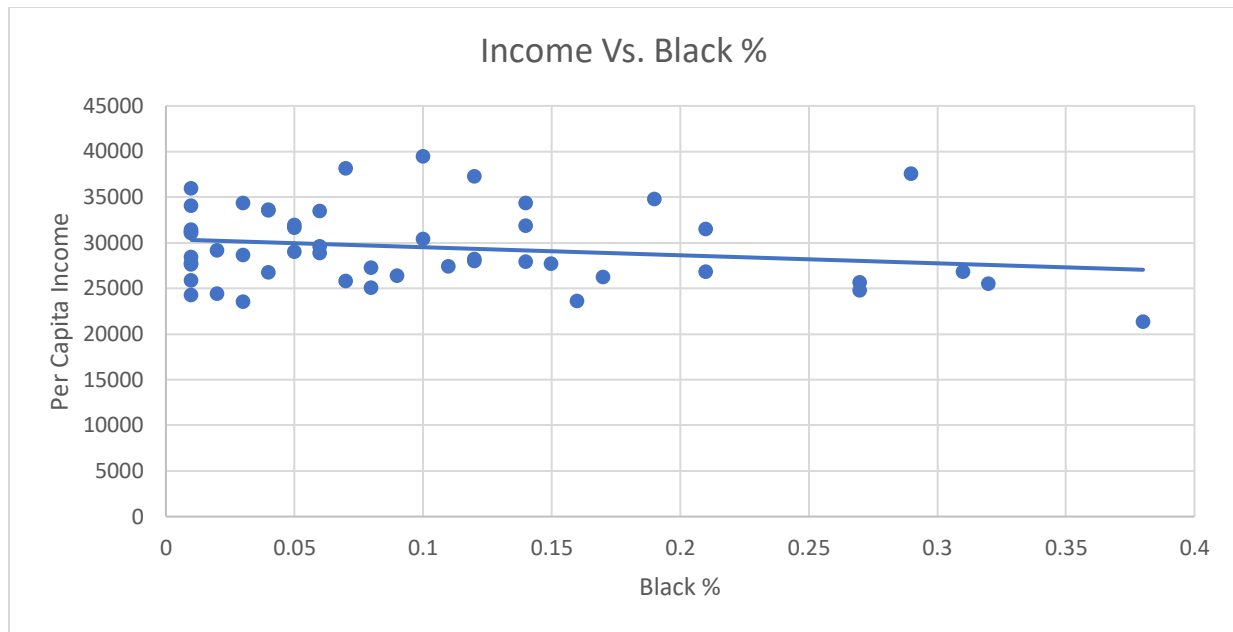
There were a few surprises in the results from the regressions. The coefficient for the percentage of state population that was black was positive for every model, which contradicted previous research that suggested blacks generally have lower incomes than other races. I analyzed the data further and looked at the direct relationship between income and % of

population that is black. Running a linear regression with least sum squares for the 2015 data resulted in: **Per_Capita_Income = 6045*Black + 29245.**

When looking at a graph of Per_Capita_Income vs. Black, we see that there is a significant outlier with high income and high black %, which is the District of Columbia.



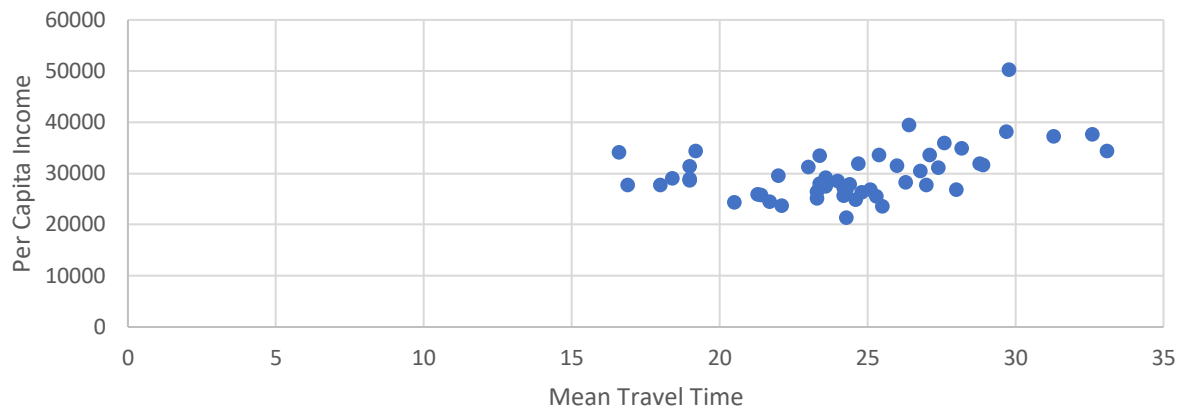
When this outlier is removed, the relationship becomes negative, with a linear model of:
Per_Capita_Income = -8820.08*Black + 30396.



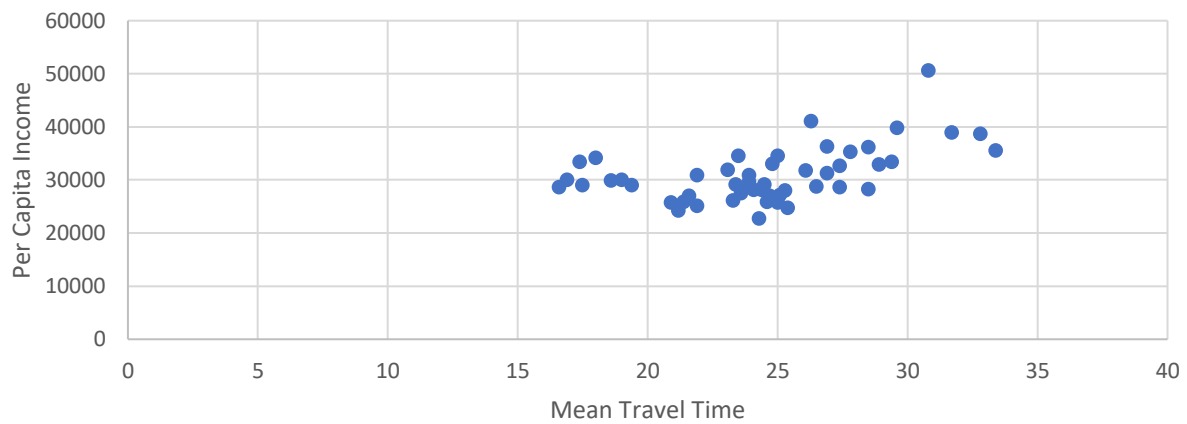
This new negative correlation is consistent with previous research, and thus I found it likely that the positive correlation seen in the initial model was due to the District of Columbia outlier.

Average commuting times did not have a very significant effect on the model according to the respective p-values from each year. This was especially the case in the 2015 model, as the variable for mean commuting time (or mean travel time) had a p-value of 0.545386. However, the p-value did drop substantially in 2016 and 2017 to 0.08429 and 0.057528. When examining the data, there doesn't appear to be any outliers.

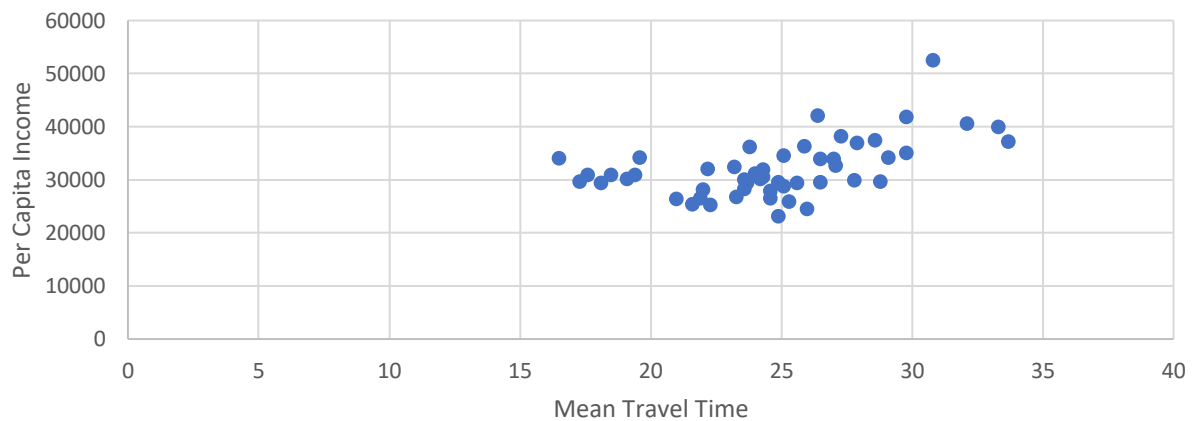
Commute Time vs Income (2015)



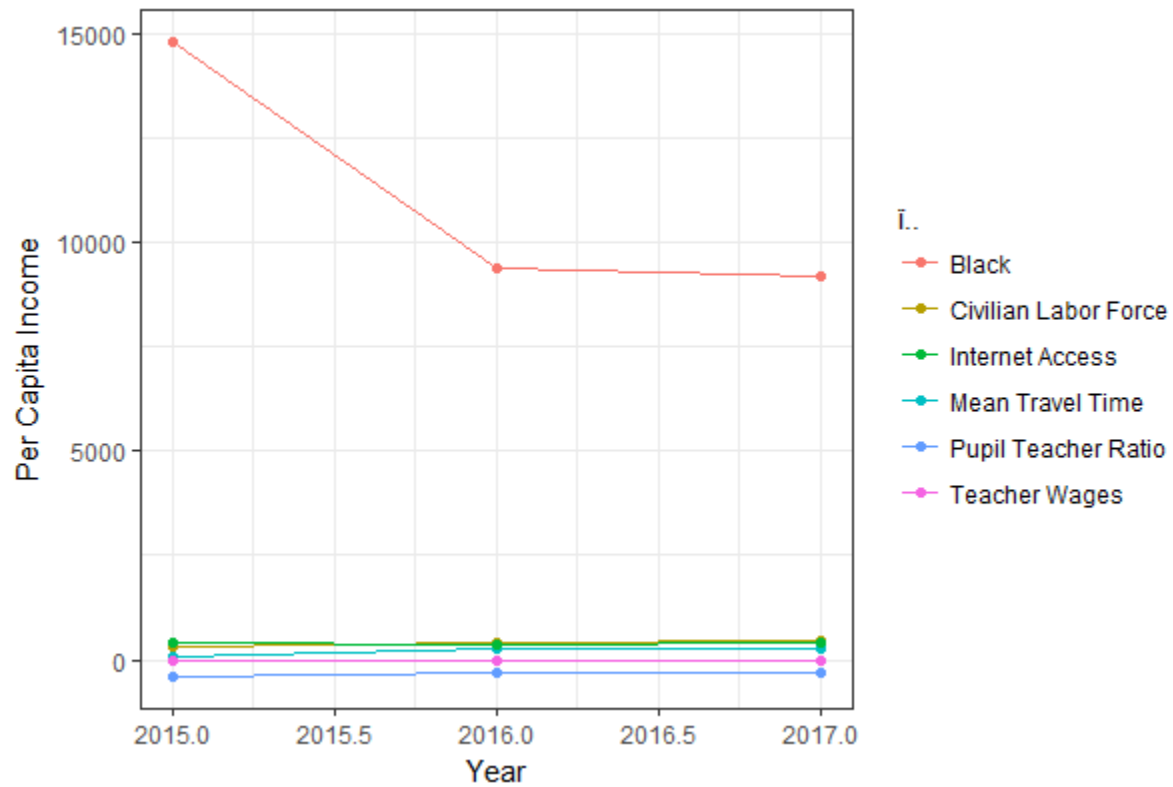
Commuting vs. Income (2016)



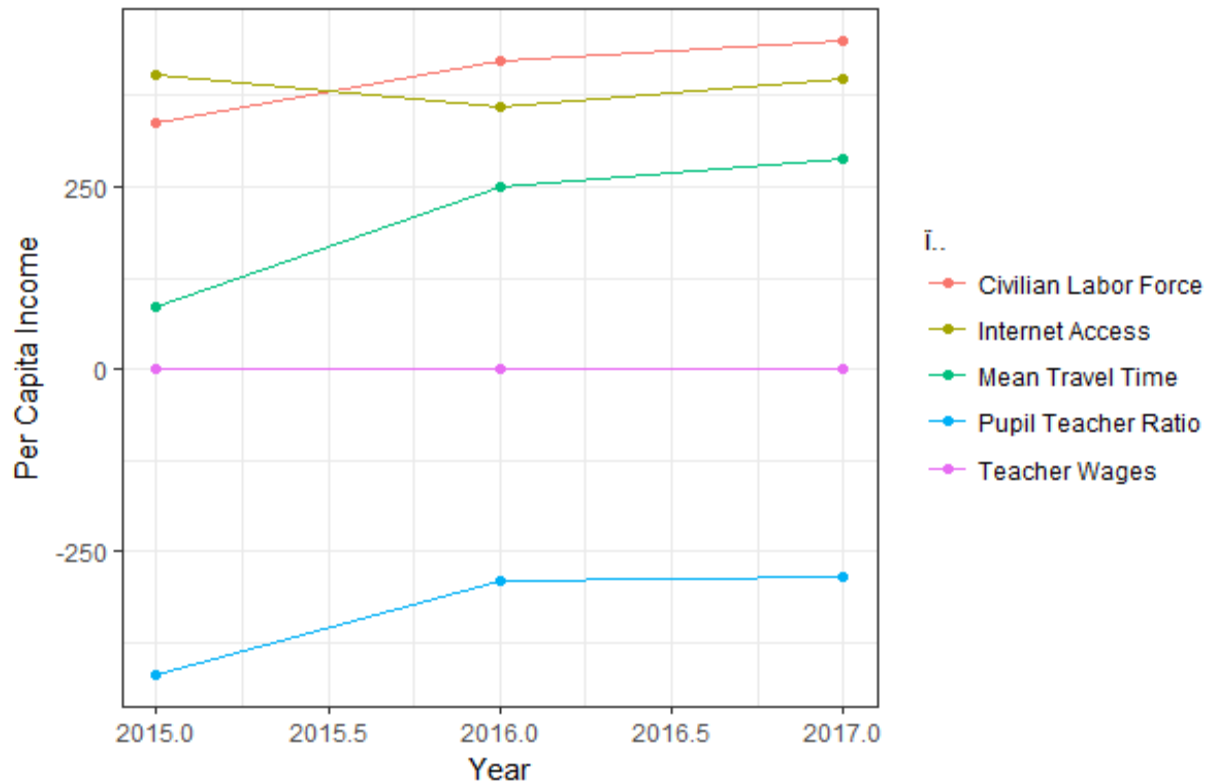
Commuting vs. Income (2017)



When analyzing how the coefficients of each variable fluctuated throughout the time period of the study, it is initially very clear the size of the effect of black % on income decreased considerably in the first year.



In order to better examine the trends of the other variables, I removed Black % from the plot for better scale.



Both Pupil/Teacher Ratio and Mean Travel time experienced large increases in effect size, while the other variables did not undergo any substantial change in effect size. It is also interesting to note that although Teacher Wages had a statistically significant effect on income per capita, it did not seem to have a practical effect in terms of coefficient size.

Discussion:

The results of each linear model allow us to conclude that the percentage of a population in the labor force, the percentage of population that is black, total population with internet access, pupil/teacher ratio and average teacher salaries all have statistically significant effects on individual income per capita.

However, when closely examining the results and coefficients, there are some inconsistencies to account for. The fact that the black percentage variable had a positive coefficient for each year was likely due to an outlier in that statistic, and this could have led to other inaccuracies in the model.

The extremely small coefficient value for teacher wages was another area of concern, yet according to the 1992 study by David Card and Alan Krueger a substantial 30% increase in teacher pay correlated with increased average future earnings of 1.34%. I found these earnings increase to be very minor, and thus the results from my model may have actually reinforced the findings in the study by David Card and Alan Krueger.

Finally, there was an interesting pattern where half of the independent variables underwent dramatic changes to their coefficients from 2015 to 2016. Although there are likely a multitude of economic and social factors that could have caused these fluctuations, I would speculate that the economic shift that occurred in 2015 could be partly responsible, which is when the number of Americans in the middle-class was outnumbered by those in the upper and lower classes for the first time in four decades (Pew Research 1). This would highlight the increasing income inequality in the US at the time, and an example of what effect this could have had would be that the increase in income inequality could have negatively affected the performance of teachers who are normally on the lower end of the income scale, thereby reducing the effect that lowered pupil/teacher rates would have on income.

Works Cited:

- Bobbitt-Zeher, D. (2007). The Gender Income Gap and the Role of Education. *Sociology of Education*, 80(1), 1–22. <https://doi.org/10.1177/003804070708000101>
- Card, D., & Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, 100(1), 1–40. <https://doi.org/10.1086/261805>
- Carra, G., Mulalic, I., Fosgerau, M., & Barthelemy, M. (2016). Modelling the relation between income and commuting distance. *Journal of the Royal Society Interface*, 13(119), 20160306. <https://doi.org/10.1098/rsif.2016.0306>
- PEW RESEARCH CENTER. (2016, June 27). *Demographic trends and economic well-being*. Pew Research Center's Social & Demographic Trends Project. <https://www.pewresearch.org/social-trends/2016/06/27/1-demographic-trends-and-economic-well-being/>
- Sachs, J. D. (2003, February 10). *Institutions Don't Rule: Direct Effects of Geography on Per Capita Income*. Wwww.nber.org. <https://www.nber.org/papers/w9490>