Hayley Brown
23 March 2018

Black Educational Attainment in Large Majority Black Cities

Educational attainment among blacks is often reported in relation to whites, but this can obscure how other factors influence educational and economic outcomes for blacks in relation to each other. To this end, my objective is to examine where blacks in large, majority-black (>50.1%) cities are the most educated. Specifically, I examine the percentage of black residents with 4-year, graduate and professional degrees in majority-black cities with populations of 100,000[1] or more. According to the most recent ACS Demographic and Housing Estimates, there are 21 cities in the United States that fit the criteria for analysis (see Appendix A for rankings). The majority-black city with the highest percentage of black residents with at least a bachelor's degree is Washington, DC, while the city with the lowest percentage of such residents is Cleveland, Ohio. A geographical representation of these findings is mapped in Appendix B. In order to ascertain what may differentiate cities with more highly educated black residents from cities with fewer highly educated black residents, I consider four additional factors: median household income, per-capita income, supplemental nutrition assistance program (SNAP) usage, and high school graduation rate. The first three roughly relate to income level, while the latter speaks directly to lower levels of robustness in the educational system. I use R to run linear regression analyses for each of these variables separately, and a multiple regression analysis for these variables in conjunction with one another; my code can be found in Appendix D.

Some of my findings are as follows:

- Among the cities examined, income serves as a key differentiating factor. Specifically, cities where a higher percentage of black residents possess at least a bachelor's degree tend to have higher per capita (Appendix C, Figure 1) and median household incomes (Appendix C, Figure 2). The correlation may reflect that college graduates typically earn more money. The direction of causality is unclear (and may be circular) – highly educated parents with more money to spend on their children may be better able to invest in those children's education. More granular, temporal data would be useful for future inquiry in this regard.
- High school graduation rates also correlate significantly with higher levels of post-secondary educational attainment (Appendix C, Figure 3). This is not surprising, since one generally must graduate from high school before embarking on a university degree. However, the correlation between the two suggests that part of the problem may need to be tackled at the K-12 level.
- I run a multiple regression analysis, with the percentage of black residents who possess a bachelor's degree or higher as my dependent variable, and high school graduation rate, per capita income, median household income, and SNAP usage rate

[1] Unfortunately, Flint, MI missed this cutoff by a mere few hundred people (total population of 99, 802), but rules are rules.

as my independent variables. The results indicate that 66.65% (or, adjusted based on the number of samples, 58.31%) of the variation in the cities' percentage of college-educated black residents can be explained in terms of those four independent variables. The low p-value indicates that these results are statistically significant.

- Other spatial variables that may be worth looking into include zoning practices, pollutants, housing prices, and/or foreclosure rates. These would be especially amenable to spatial econometric analyses, such as fixed-effect spatial lag models. I was unfortunately unable to gather sufficient data on these variables in the time allotted, but I strongly suspect that they would be illuminating.
- How much of the highly skilled workforce is imported vs. home-grown? Are the cities with the highest educational attainment rates attracting black college graduates from other places? Importing a strong workforce will provide a helpful initial boost, but a sustainable long-term strategy must include local children and institutions – long-term infrastructure for success.

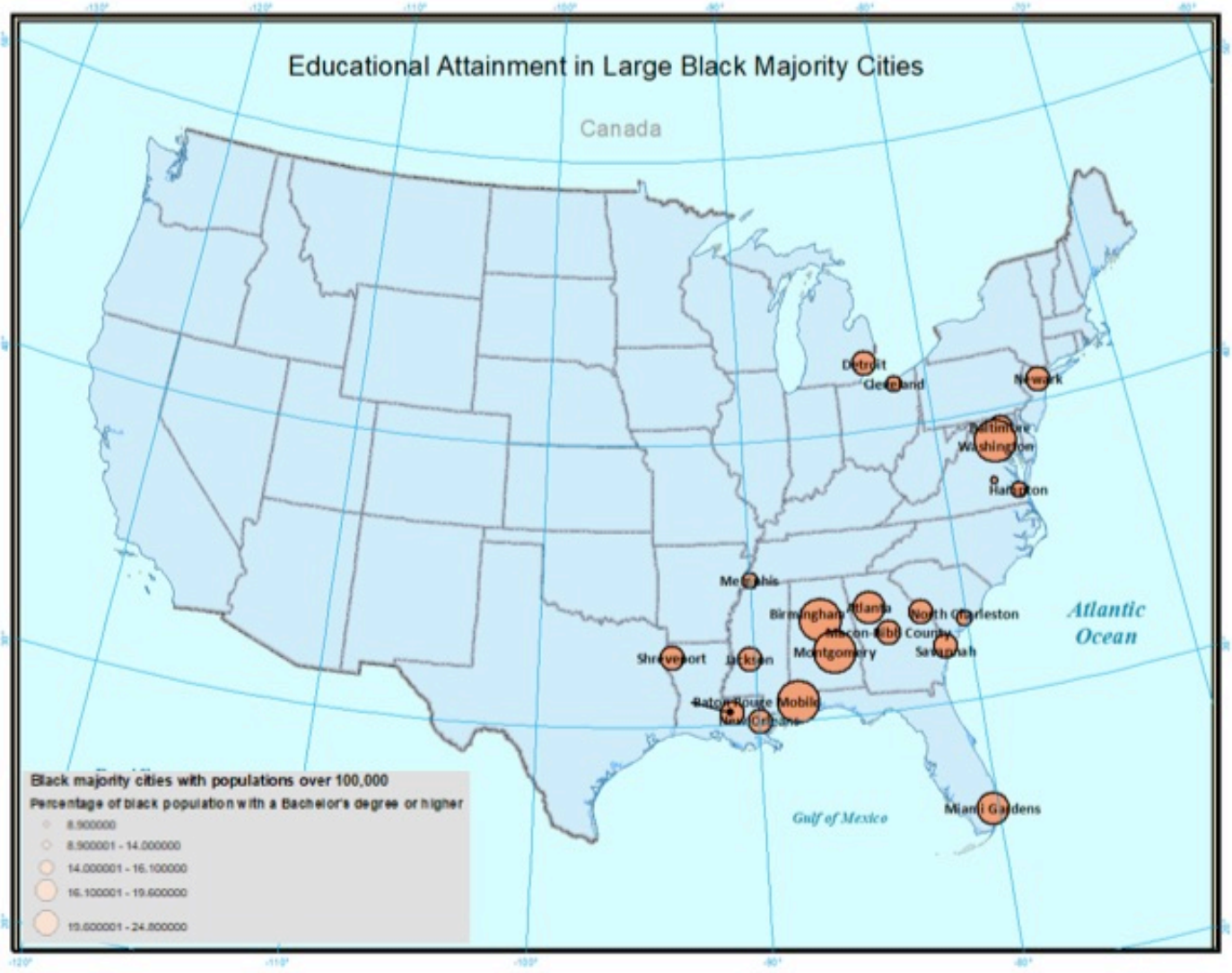
My findings indicate that the combination of various measures of income and high school graduation are strong predictors of the possession of bachelor's degrees among black residents in large black-majority cities. The strong explanatory power of these variables suggest that policies which increase economic opportunity will correlate with increases in the number of bachelor's degree holders in these places. However, the presence of high-paying jobs that reward higher education is not enough; it must correspond with investment in the local workforce, including improvements in K-12 and other programs that encourage community achievement. It is also important for these types of programs to retain links to local stakeholders, so that the residents who make these cities thrive can in turn reap the benefits of their successes. That is, success should not mean gentrification that displaces residents.

Appendix A: Majority Black Cities with Populations over 100,00, ranked by percentage of population 25+ with a bachelor's degree or higher

Rank	Geography	Total Population	Black Population	% Black Population	Percent of Black population over 25 with Bachelor's degree or higher
1	Washington city, District of Columbia	622,454	400,587	64.4%	24.8%
2	Atlanta city, Georgia	390,584	209,931	53.7%	22.9%
3	Hampton city, Virginia	112,021	84,151	75.1%	22.4%
4	Montgomery city, Alabama	647,484	324,552	50.1%	21.7%
5	Jackson city, Mississippi	194,669	99,749	51.2%	19.6%
6	New Orleans city, Louisiana	212,211	155,377	73.2%	18.1%
7	Savannah city, Georgia	229,186	128,781	56.2%	16.1%
8	Miami Gardens city, Florida	448,901	242,006	53.9%	15.6%
9	Memphis city, Tennessee	137,081	71,407	52.1%	15.4%
10	Birmingham city, Alabama	690,074	562,887	81.6%	15.3%
11	Baton Rouge city, Louisiana	213,735	108,694	50.9%	15.0%
12	Augusta-Richmond County consolidated government (balance), Georgia	376,738	227,444	60.4%	14.8%
13	Baltimore city, Maryland	200,015	113,100	56.5%	14.8%
14	Macon-Bibb County, Georgia	279,793	146,436	52.3%	14.6%
15	Richmond city, Virginia	104,146	52,419	50.3%	14.6%
16	Mobile city, Alabama	657,167	419,729	63.9%	14.4%
17	North Charleston city, South Carolina	142,919	78,906	55.2%	14.0%
18	Shreveport city, Louisiana	202,967	120,855	59.5%	13.6%
19	Newark city, New Jersey	173,212	139,778	80.7%	13.0%

20	Detroit city, Michigan	196,635	112,214	57.1%	12.1%
21	Cleveland city, Ohio	154,608	83,297	53.9%	8.9%

Appendix B: Map of Educational Attainment Discrepancies in Large Black Majority Cities



Appendix C
Figure 1

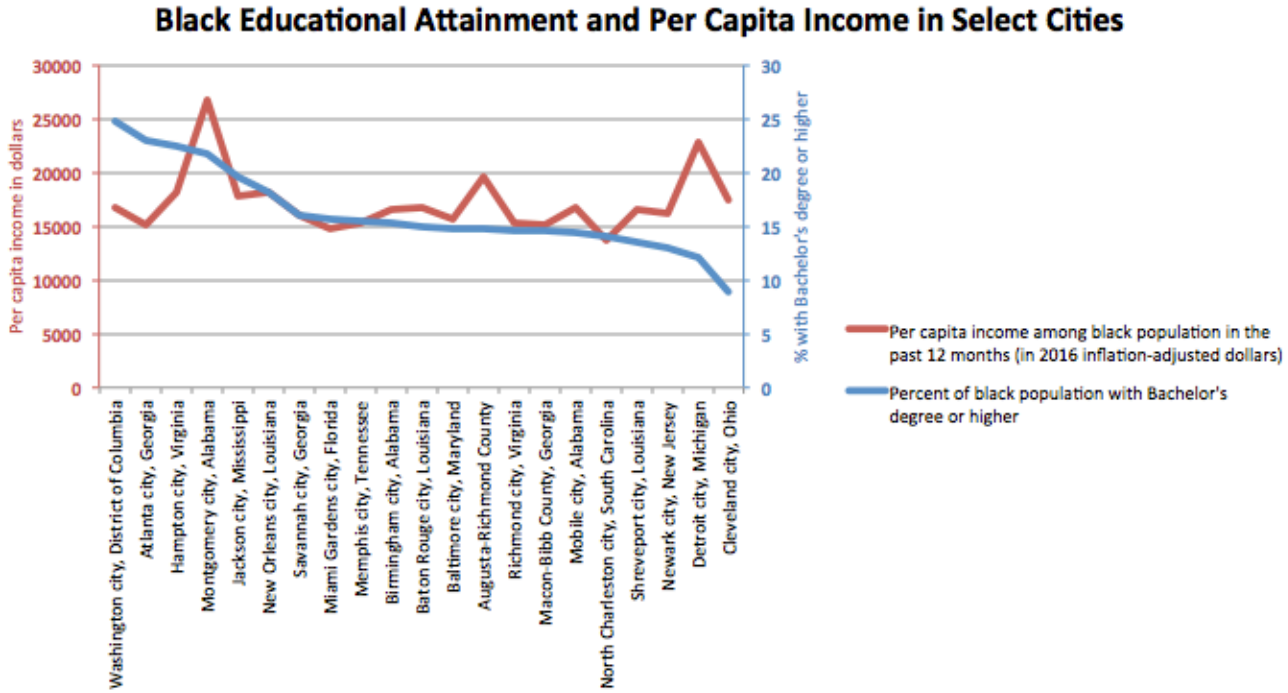


Figure 2

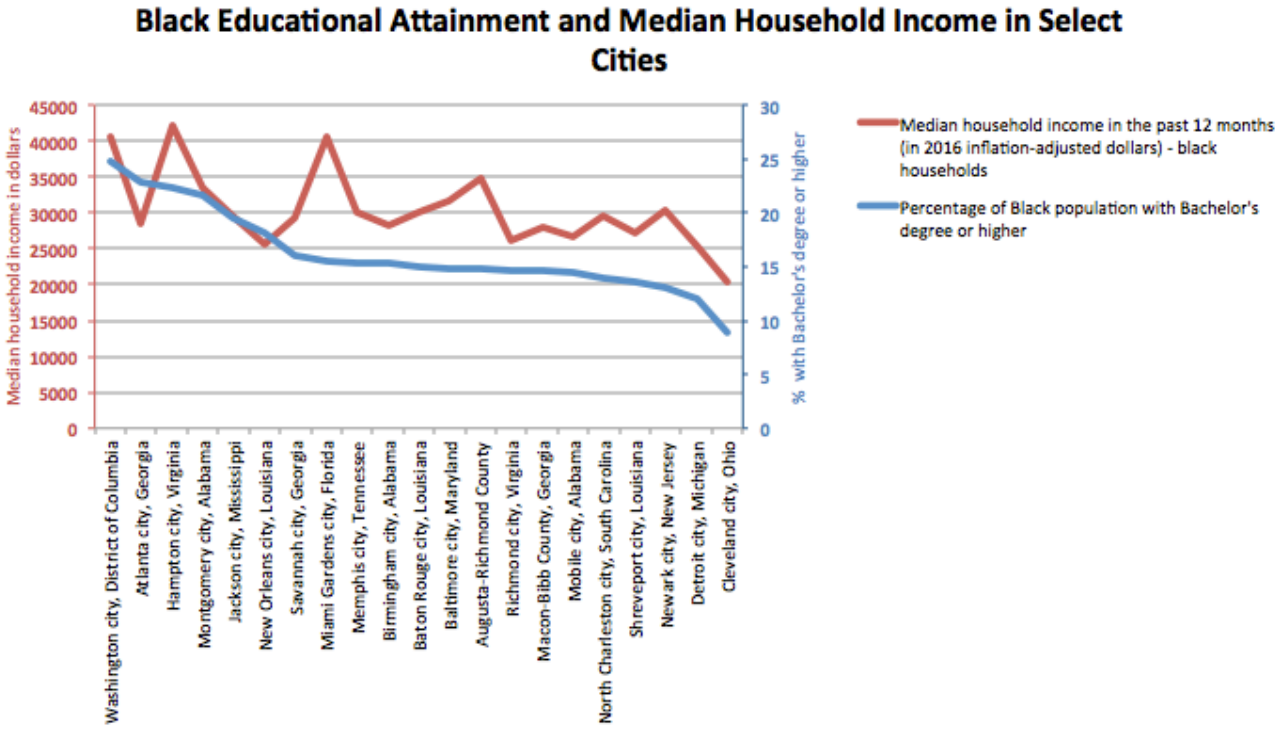
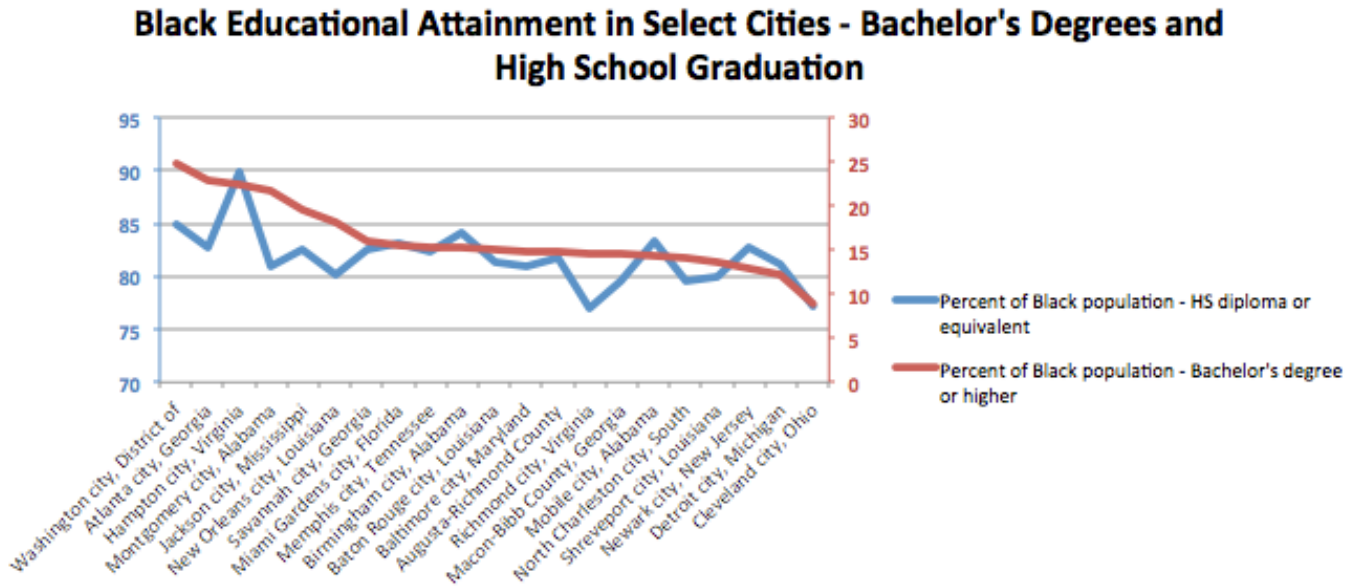


Figure 3



Appendix D: R Code

```
# Read the data into a data frame in R
brookings <- read.csv ("~/Documents/brookings-related/brookingsforR2.csv")
View(brookings)


# Rename things to make them more accessible
bach <- brookings$percentblackbach
hs <- brookings$percentblackhs
medincome <-brookings$medincome
percapincome <- brookings$percapincome
SNAP <- brookings$percentSNAP

# Run a regression using lm( ), where bach (percentage of black bachelor's
degree earners) is the dependent variable, hs (percentage of black high
school graduates) is the independent variable
lin.mod<-lm(bach~hs)
lin.mod

Call:
lm(formula = bach ~ hs)

Coefficients:
(Intercept)           hs
   -52.9863       0.8467
summary(lin.mod)
Call:
lm(formula = bach ~ hs)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-4.0375 -1.7762 -0.7682  2.4735  6.1866

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -52.9863    21.6435  -2.448   0.0242 *
hs            0.8467     0.2645   3.202   0.0047 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.256 on 19 degrees of freedom
Multiple R-squared:  0.3504,  Adjusted R-squared:  0.3163
F-statistic: 10.25 on 1 and 19 DF,  p-value: 0.004697

# Reject null hypothesis; there’s a correlation here!

# Run a regression using lm( ), where bach (percentage of black bachelor’s
degree earners) is the dependent variable, medincome (median income) is the
independent variable

lin.mod<-lm(bach~medincome)
lin.mod

Call:
lm(formula = bach ~ medincome)

Coefficients:
(Intercept)    medincome
   2.622399     0.000449


summary(lin.mod)
Call:
lm(formula = bach ~ medincome)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2575 -1.9186 -0.6108  0.8865  7.4966

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.6223988  4.0612372   0.646  0.52619
medincome   0.0004490  0.0001316   3.411  0.00293 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.181 on 19 degrees of freedom
Multiple R-squared:  0.3798,  Adjusted R-squared:  0.3472
```

```
F-statistic: 11.64 on 1 and 19 DF,  p-value: 0.002931
# Again, kids, we've got some significance here!

# Run a regression using lm( ), where bach (percentage of black bachelor's
degree earners) is the dependent variable, percapincome (per capita income)
is the independent variable

lin.mod<-lm(bach~percapincome)
lin.mod

Call:
lm(formula = bach ~ percapincome)

Coefficients:
 (Intercept)  percapincome
   -0.741702      0.000989

summary(lin.mod)

Call:
lm(formula = bach ~ percapincome)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8792 -1.6270 -0.5268  0.6804  5.6400

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.7417025  3.6715370  -0.202 0.842053
percapincome  0.0009890  0.0002106   4.697 0.000157 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.748 on 19 degrees of freedom
Multiple R-squared:  0.5373,  Adjusted R-squared:  0.5129
F-statistic: 22.06 on 1 and 19 DF,  p-value: 0.000157

# Run a regression using lm( ), where bach (percentage of black bachelor's
degree earners) is the dependent variable, SNAP (percentage of resident's
receiving food assistance) is the independent variable

lin.mod<-lm(bach~SNAP)
lin.mod

Call:
lm(formula = bach ~ SNAP)

Coefficients:
(Intercept)         SNAP
```

```
    24.4284      -0.3516

summary(lin.mod)

Call:
lm(formula = bach ~ SNAP)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0426 -2.1504 -0.5221  2.6478  5.6455

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 24.42840    2.41073  10.133 4.25e-09 ***
SNAP        -0.35159    0.09963  -3.529  0.00224 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.14 on 19 degrees of freedom
Multiple R-squared:  0.396,   Adjusted R-squared:  0.3642
F-statistic: 12.45 on 1 and 19 DF,  p-value: 0.002242

# MULTIPLE REGRESSION
# Run a multiple regression with percentage of black bachelor’s degree
earners (bach) as the dependent variable, and high school graduation (hs),
per capita income (percapincome), median income (medincome), and SNAP
enrollment (SNAP) as independent variables.

lin.mod2<-lm(bach~hs+percapincome+medincome+SNAP)
lin.mod2<-lm(bach~hs+percapincome+medincome+SNAP)
lin.mod2

Call:
lm(formula = bach ~ hs + percapincome + medincome + SNAP)

Coefficients:
 (Intercept)            hs  percapincome     medincome          SNAP
  -1.195e+01     2.777e-01     5.885e-04     3.467e-06    -2.033e-01

summary(lin.mod2)

Call:
lm(formula = bach ~ hs + percapincome + medincome + SNAP)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4655 -1.5746 -0.7424  1.0193  4.6083

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  -1.195e+01  2.206e+01  -0.542   0.5954
hs            2.777e-01  3.024e-01   0.918   0.3721
percapincome  5.885e-04  3.528e-04   1.668   0.1147
medincome     3.467e-06  1.999e-04   0.017   0.9864
SNAP         -2.033e-01  9.162e-02  -2.219   0.0413 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.542 on 16 degrees of freedom
Multiple R-squared:  0.6665,  Adjusted R-squared:  0.5831
F-statistic: 7.993 on 4 and 16 DF,  p-value: 0.0009697
```

The results indicate that 66.65% (or, adjusted based on the number of x values, 58.31%) of the variation can be explained in terms of the listed variable. Based on these results, I reject the null hypothesis.