Rent Burden in Massachusetts

A Quantitative Analysis

Emma Colley
SES 5215

Research Questions and Hypothesis

This study aims to examine rent burden in Massachusetts. Having worked on and lived in affordable housing, this topic is very important to me. I believe planners can and must prioritize solutions to rent buden.

In order to study this, I will compare **Gross Rent as a Percentage of Household Income** to several other variables, which we will list shortly. During my study, the variables that I examined most closely were **household income, race and educational attainment.**

Research Questions and Hypothesis

Does a person's household income, race, or educational attainment have a relationship to the percentage of their income they pay toward rent each month?

My hypotheses:

People in minority racial groups (people who do not identify as white) have a higher GRPIP (gross rent as a percentage of household income.)

People with a lower level of education have a higher GRPIP (gross rent as a percentage of household income.)

Lower-income households have higher GRPIP (gross rent as a percentage of household income.)

Defining Sample Set

The unit of analysis is **person-level data** from the **2018 American Community Survey.**

The definition of the sample population is people **over in the age of 18** in **rental households** in the state of Massachusetts.

For the purposes of this exercise, I am defining a "rent-burdened" induvidual as someone who spends **30% or more of their household income on rent**. I used 30% because this is the percentage HUD uses to define cost-burdened families.

Defining Variables

These are the orginal variables I worked with from the 2018 ACS:

Continuous

Gross Rent as Percentage of Household Income (GRPIP)
Annual Household Income (HINCP)
Age of Individual (AGEP)
Number of People in Household (NP)

Categorical

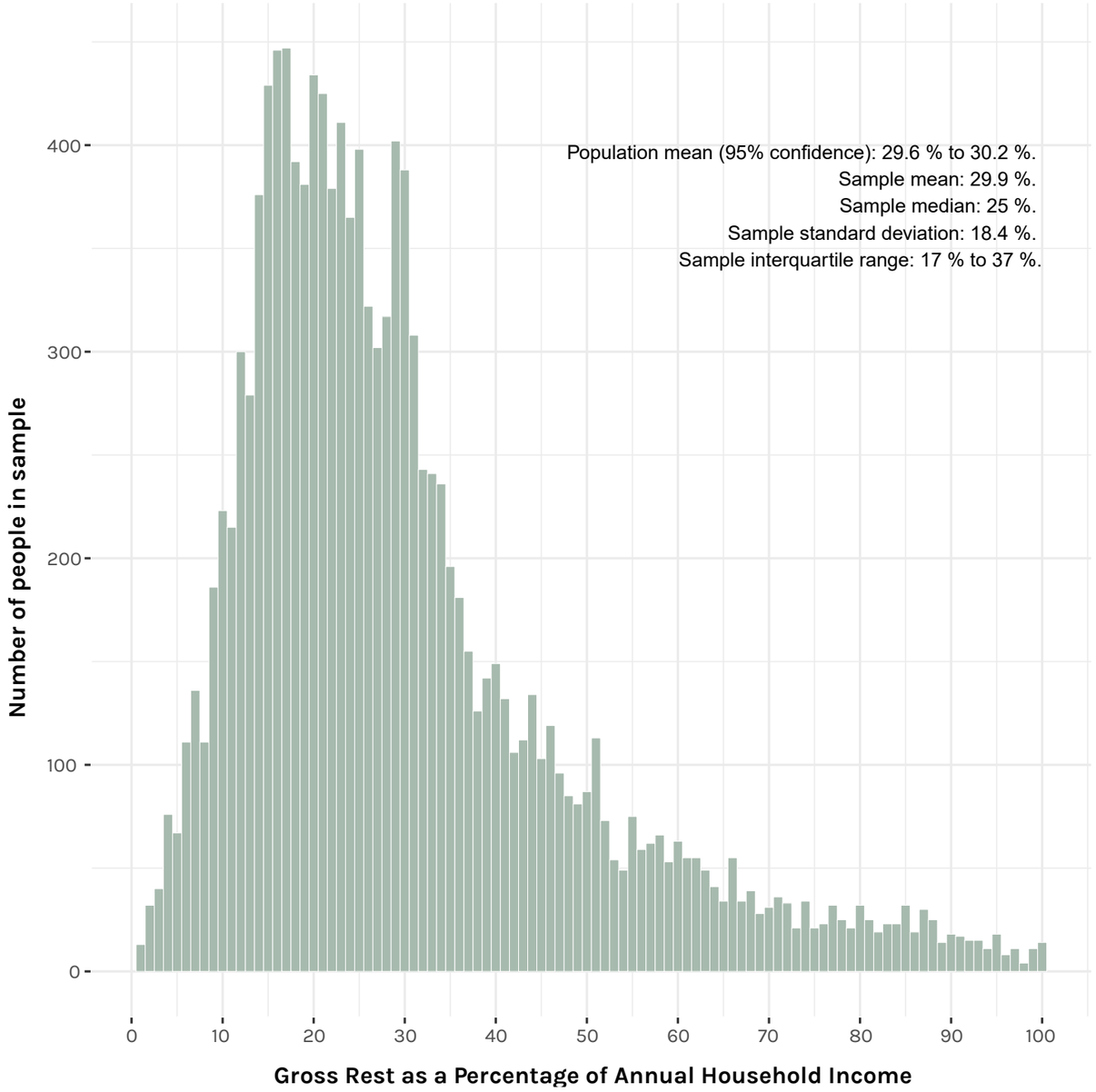
Race/Ethnicity (RAC1P)
Highest Level of Education Received (SCHL)

The following slides show the three variables I am most interested in:

Distributions and Proportions

Gross Rent as a Percentage of Household Income

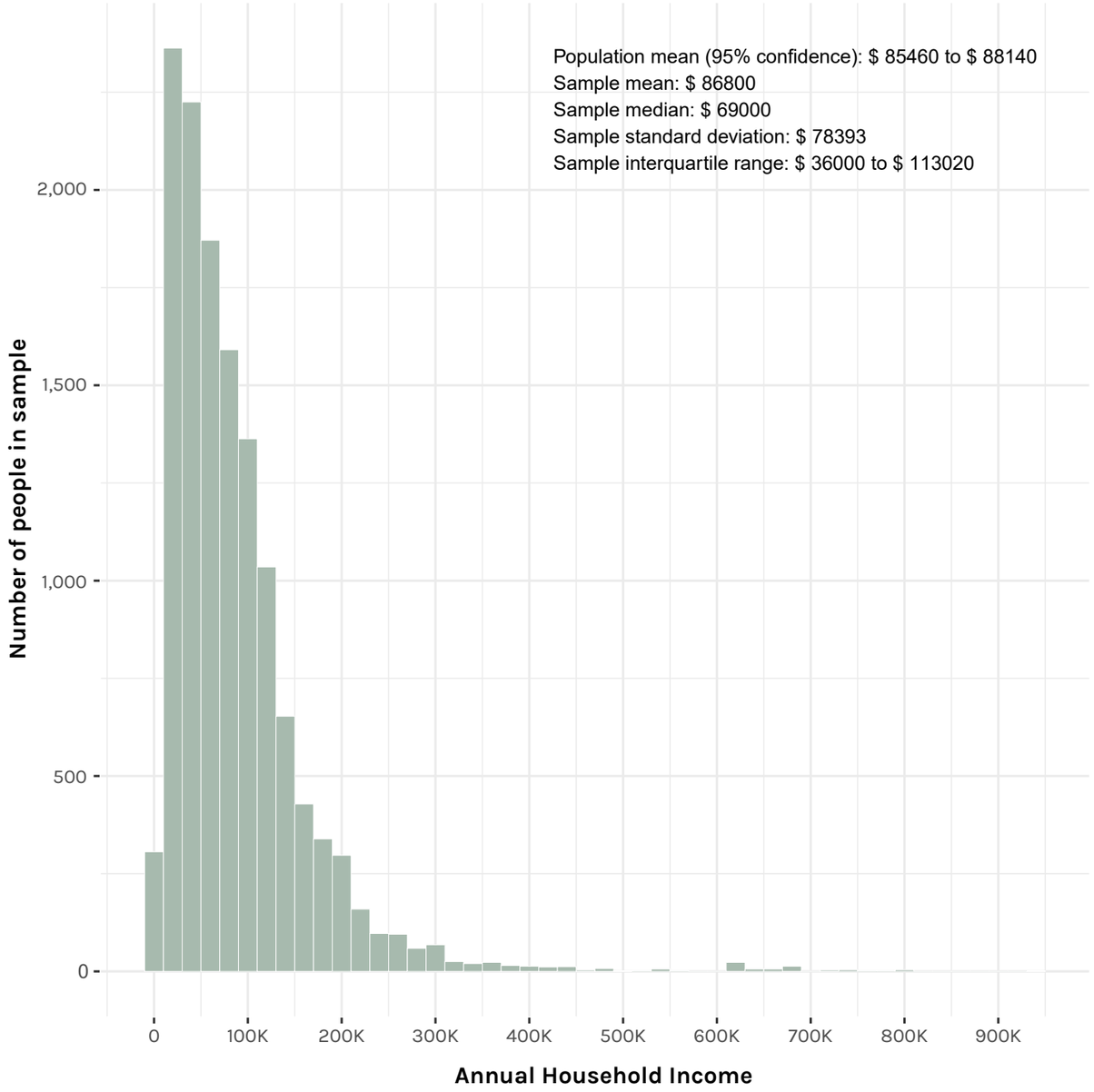
Key takeway: the population mean of about 30% tells us that, on average, renters in MA are right at the threshold of being rent burdened.



Distributions and Proportions

Annual Household Income

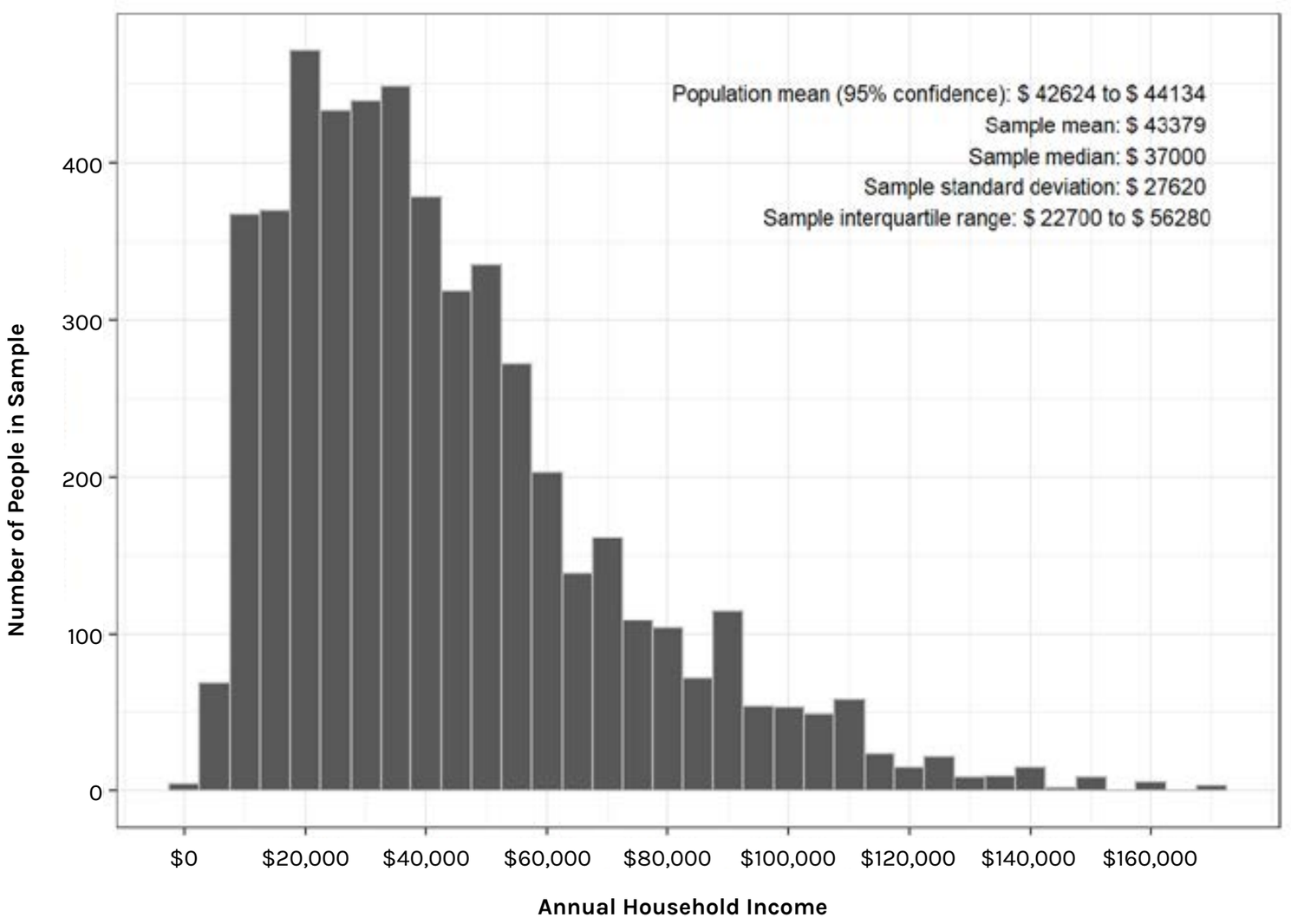
Key takeway: the population mean of $85,460 - $88,140 median of $69,000 is high compared to the median household income in the U.S. of $60,000.



Distributions and Proportions

Annual Household Income (of rent burdened people)

Key takeway: when you look at the population of only rent burdened induviduals (GRPIP > 30%),
the mean and median income decreases drastically.

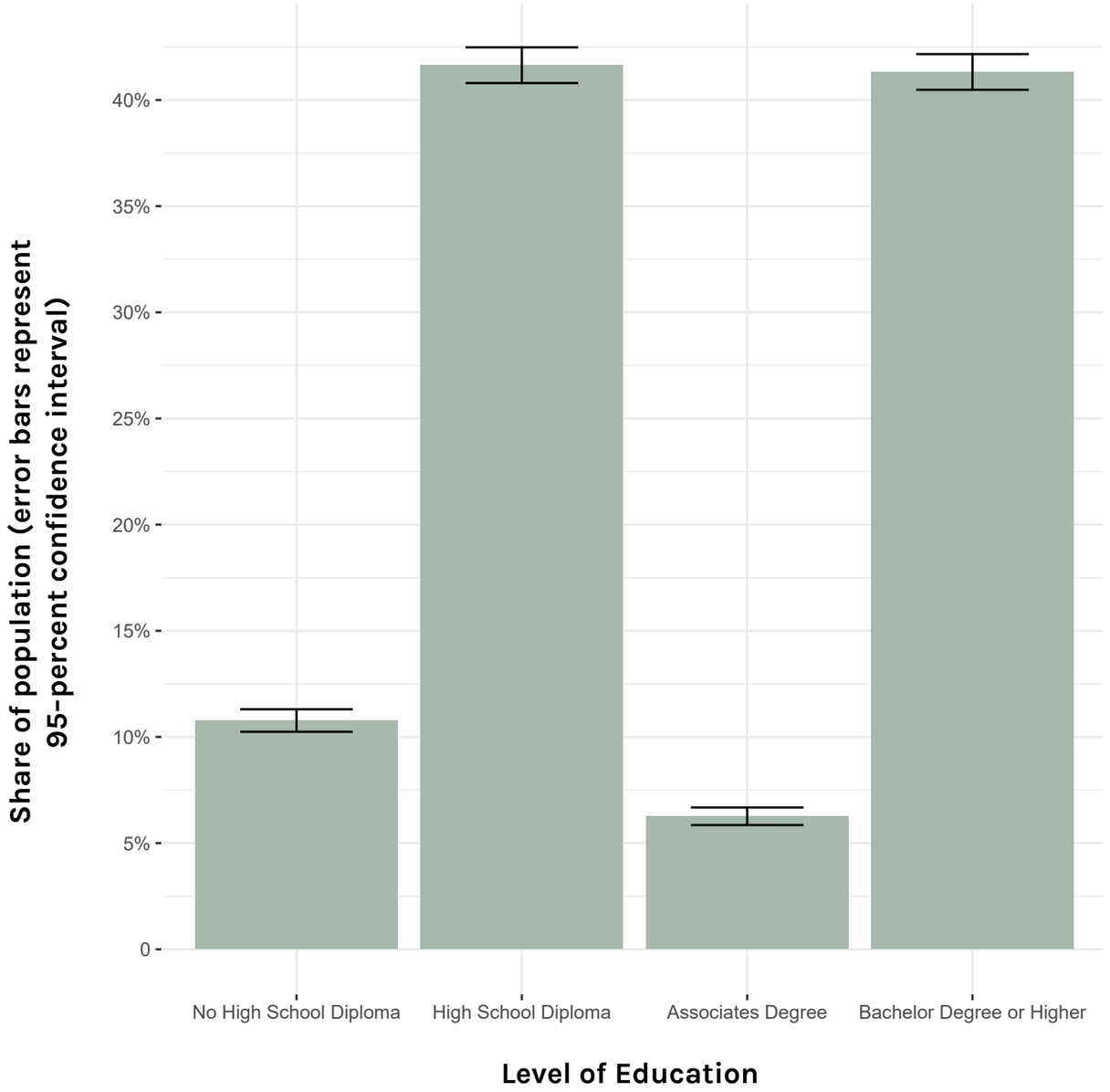


Data

Distributions and Proportions

Level of Education

Key takeway: these are broad categories I designated by coding the original ACS variable. They are probably too broad and I would want to break them down more in the future. Over 40% of the population has a Bachelor's degree or higher, which is slightly higher than the national average of 33%.

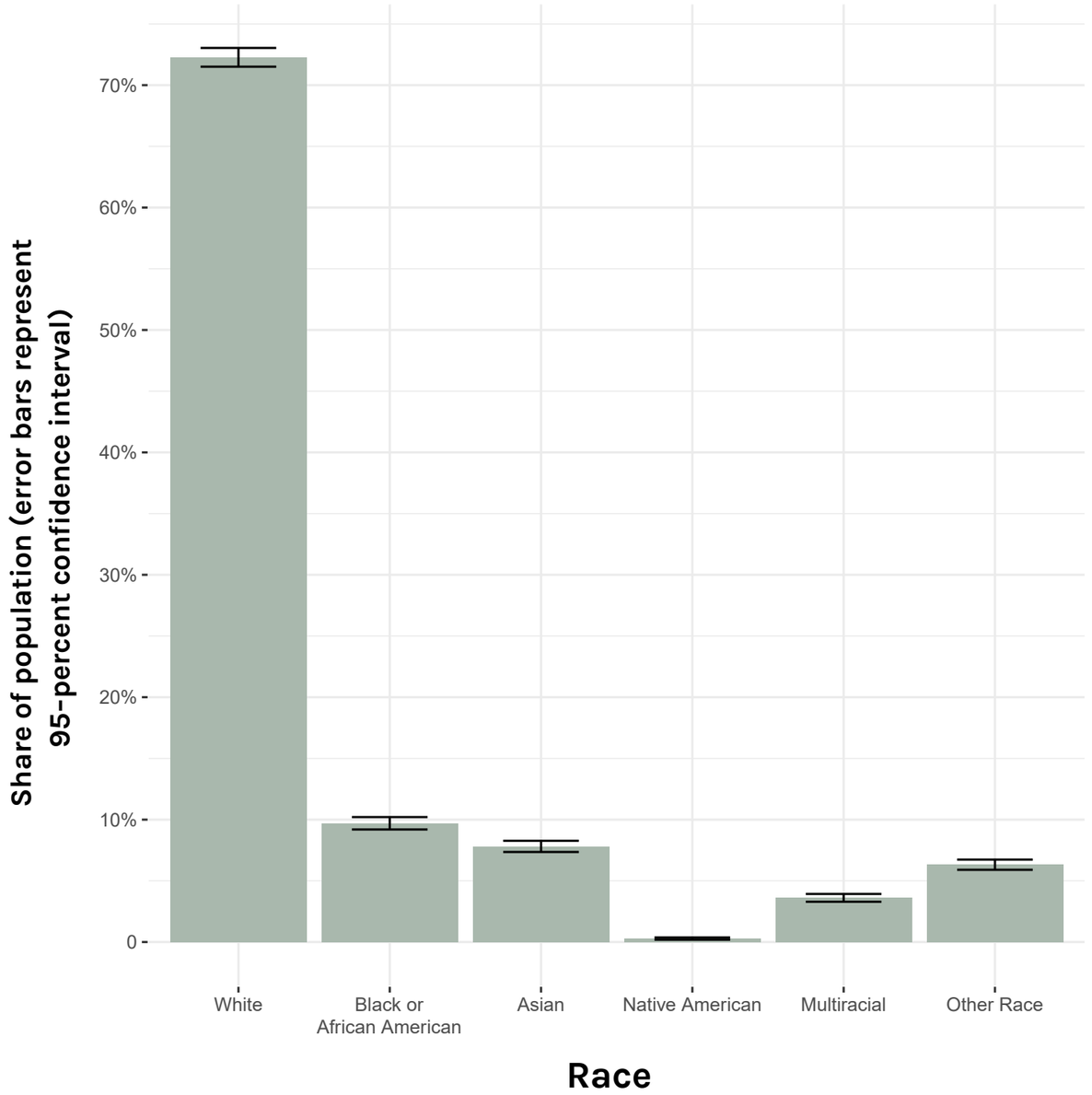


Data

Distributions and Proportions

Race

Key takeway: these are broad categories I designated by coding the original ACS variable. Over 70% of the population is white alone. I didn't realize this, but this is fairly representative of Massachusetts as a whole. Census.gov reports 80% of Massachusetts as white alone.



01 Introduction

02 Data

03 **Methods & Results**

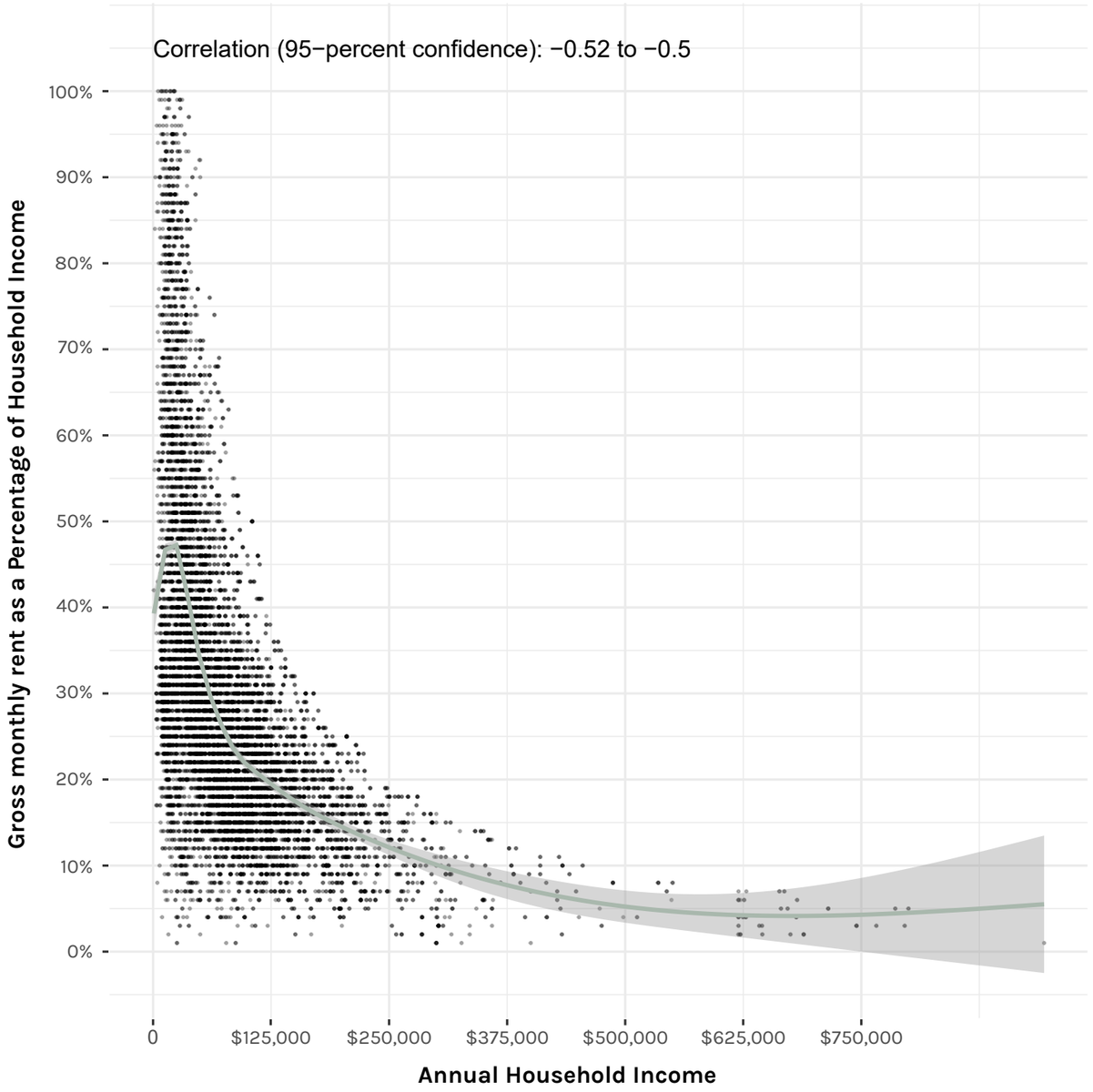
04 Discussion

05 Conclusion

Pearson's Test

Is there a correlation between Household Income and Gross Rent as a Percentage of Household Income?

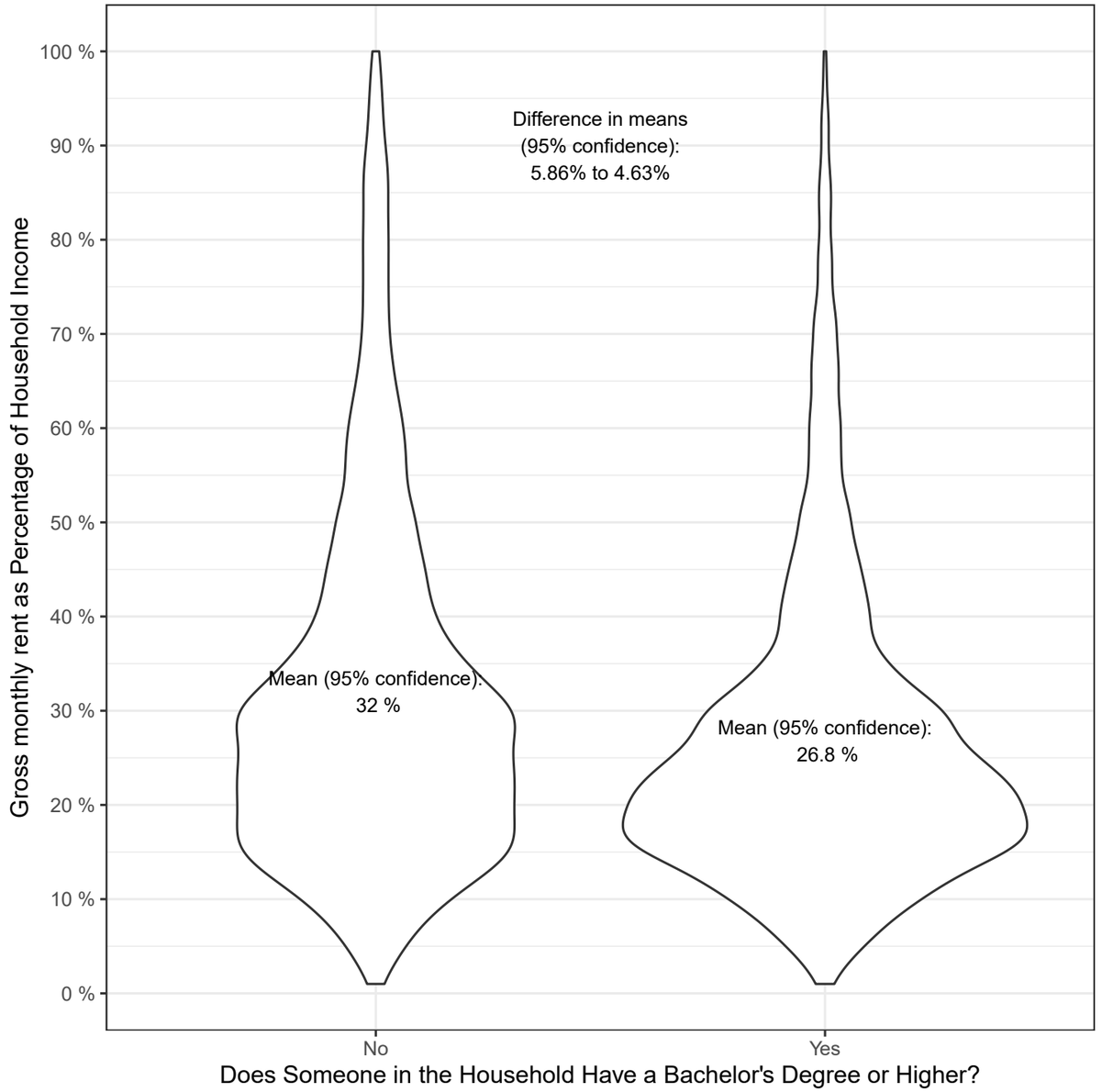
This correlation is only moderate (around 0.5) but statistically significant. There is a negative correlation at a 95% confidence level.



Two-Sample t-tests

Are people with a Bachelor's Degree less likely to live in rent burdened households?

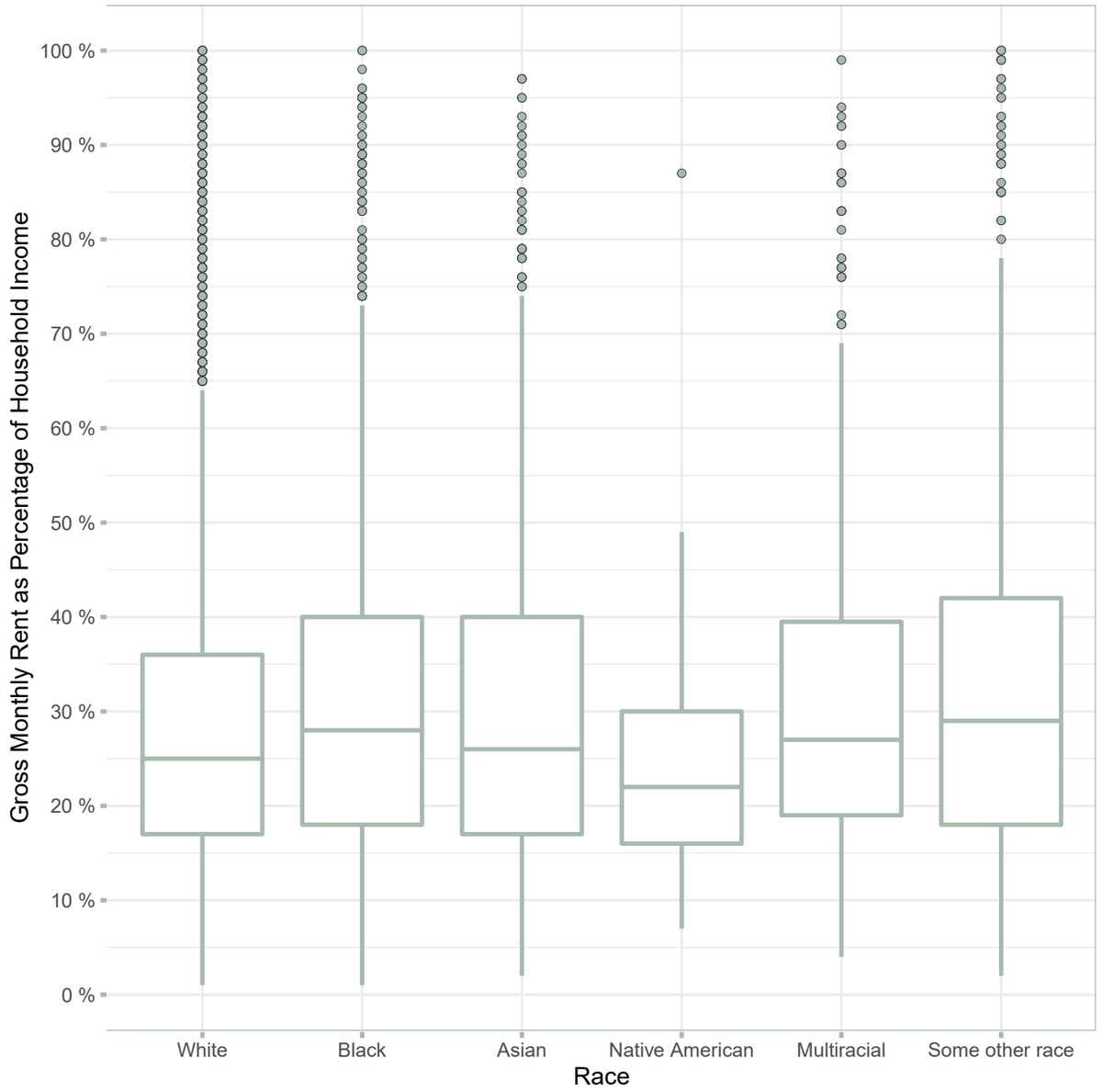
There is a significant relationship between these two variables. We can be 95% confident that people in this sample set with a Bachelor's degree or higher live in households that pay 5% - 6% less of their income on rent than those without a Bachelor's degree. It looks like many people with a college degree live in households that pay less than 20% of their income on rent.



Tukey / ANOVA

Is there a correlation between Gross Rent as a Percentage of Household Income and Race?

There is a honestly significant relationship between these variables. Here is their distribution shown in a box box. The upper and lower quartiles of each racial category are not dramatically different. The lowest median GRPIP is seen in the Native American or Pacific Islander population, followed by the White population.



Linear Regression | First Model

My dependant variable is Gross Rent as a Percentage of Household Income.

My independant variables are Educational Attainment, Race, Age and Number of People in the Household.

The **multiple R-squared value for this model is 0.044**, which tells us that this model predicts about 4% of the variation in gross rent as a percentage of household income in this dataset.

```
##
## Call:
## lm(formula = GRPIP ~ EDU + RACE + AGEP + NP, data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.125 -11.969  -4.419   6.972  75.003
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               39.887140   0.868523  45.925  < 2e-16 ***
## EDUAssoc. Degree          -5.229010   0.797713  -6.555 5.77e-11 ***
## EDUBach Degree or Higher -8.695419   0.575215 -15.117  < 2e-16 ***
## EDUHS Diploma             -3.146645   0.547474  -5.748 9.25e-09 ***
## RACEAsian                  3.121303   0.596580   5.232 1.70e-07 ***
## RACEBlack                  2.785634   0.542295   5.137 2.84e-07 ***
## RACEMultiracial            2.393221   0.851052   2.812  0.00493 **
## RACENative Am.            -4.890339   2.961429  -1.651  0.09869 .
## RACEOther Race             2.610134   0.670054   3.895 9.85e-05 ***
## AGEP                      -0.014633   0.009619  -1.521  0.12824
## NP                        -1.884364   0.121329 -15.531  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.97 on 13142 degrees of freedom
## Multiple R-squared:  0.04425,    Adjusted R-squared:  0.04352
## F-statistic: 60.85 on 10 and 13142 DF,  p-value: < 2.2e-16
```

Linear Regression | First Model

I am comparing **education levels** to the least amount of education: No High School Diploma. Compared to a household with no high school diploma:

A household with someone with a high school diploma pays, on average, 3.2 percentage points less on rent as a percentage of their household income.

A household with someone with an associate's degree pays, on average, 5.2 percentage points less on rent as a percentage of their household income.

A household with someone with a Bachelor's degree or higher pays, on average, 8.7 percentage points less on rent as a percentage of their household income.

These coefficients are all statistically significant.

```
##
## Call:
## lm(formula = GRPIP ~ EDU + RACE + AGEP + NP, data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.125 -11.969  -4.419   6.972  75.003
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               39.887140   0.868523  45.925  < 2e-16 ***
## EDUAssoc. Degree          -5.229010   0.797713  -6.555 5.77e-11 ***
## EDUBach Degree or Higher -8.695419   0.575215 -15.117  < 2e-16 ***
## EDUHS Diploma             -3.146645   0.547474  -5.748 9.25e-09 ***
## RACEAsian                  3.121303   0.596580   5.232 1.70e-07 ***
## RACEBlack                  2.785634   0.542295   5.137 2.84e-07 ***
## RACEMultiracial            2.393221   0.851052   2.812  0.00493 **
## RACENative Am.            -4.890339   2.961429  -1.651  0.09869 .
## RACEOther Race             2.610134   0.670054   3.895 9.85e-05 ***
## AGEP                      -0.014633   0.009619  -1.521  0.12824
## NP                        -1.884364   0.121329 -15.531  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.97 on 13142 degrees of freedom
## Multiple R-squared:  0.04425,    Adjusted R-squared:  0.04352
## F-statistic: 60.85 on 10 and 13142 DF,  p-value: < 2.2e-16
```

Linear Regression | First Model

I am comparing race to the most common race in my sample set: White alone. Compared to a White household:

A household with at least one person who is **Asian** pays, on average, **3.1 percentage points more** on rent as a percentage of household income.

A household with at least one person who is **Black or African American** pays, on average, **2.8 percentage points more** on rent as a percentage of household income.

A household with at least one person who is **multiracial** pays, on average, **2.4 percentage points more** on rent as a percentage of household income.

A household with at least one person who is **some other race** pays, on average, **2.6 percentage points more** on rent as a percentage of household income.

These coefficients are statistically significant. The Native American / Pacific Islander variable was not.

```
##
## Call:
## lm(formula = GRPIP ~ EDU + RACE + AGEP + NP, data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.125 -11.969  -4.419   6.972  75.003
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               39.887140   0.868523  45.925  < 2e-16 ***
## EDUAssoc. Degree          -5.229010   0.797713  -6.555 5.77e-11 ***
## EDUBach Degree or Higher -8.695419   0.575215 -15.117  < 2e-16 ***
## EDUHS Diploma             -3.146645   0.547474  -5.748 9.25e-09 ***
## RACEAsian                  3.121303   0.596580   5.232 1.70e-07 ***
## RACEBlack                  2.785634   0.542295   5.137 2.84e-07 ***
## RACEMultiracial            2.393221   0.851052   2.812  0.00493 **
## RACENative Am.            -4.890339   2.961429  -1.651  0.09869 .
## RACEOther Race             2.610134   0.670054   3.895 9.85e-05 ***
## AGEP                      -0.014633   0.009619  -1.521  0.12824
## NP                        -1.884364   0.121329 -15.531  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.97 on 13142 degrees of freedom
## Multiple R-squared:  0.04425,    Adjusted R-squared:  0.04352
## F-statistic: 60.85 on 10 and 13142 DF,  p-value: < 2.2e-16
```

Linear Regression | Best Fit

I was able to get a better model fit by doing three things:

1. Added Household Income (HINCP) to the regression.

2. Converted Age from a continuous variable to a categorical variable. In this case, I have two groups: 30 years old or younger and over 30 years old.

3. Added an interaction term between Race and Household Income.

I am hypothesizing that Race has an effect on Household Income when considering how much of their income one pays on rent.

The R-squared value increased from 0.04 to 0.27, which tells us that this model predicts about 27% of the variation in gross rent as a percentage of household income in this dataset.

```
##
## Call:
## lm(formula = GRPIP ~ HINCP + EDU + RACE + age_cat + NP + HINCP:RACE,
##     data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.857  -9.988  -3.553   5.581  87.821
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.996e+01  5.277e-01  75.729  < 2e-16 ***
## HINCP                      -1.217e-04  2.241e-06 -54.306  < 2e-16 ***
## EDUAssoc. Degree           -1.357e+00  6.972e-01  -1.947 0.051615 .
## EDUBach Degree or Higher    5.958e-01  5.117e-01   1.164 0.244333
## EDUHS Diploma              -9.521e-01  4.762e-01  -1.999 0.045575 *
## RACEAsian                   1.632e+00  7.672e-01   2.127 0.033441 *
## RACEBlack                   4.566e-01  6.511e-01   0.701 0.483162
## RACEMultiracial             4.553e+00  1.173e+00   3.882 0.000104 ***
## RACENative Am.             -3.341e+00  4.463e+00  -0.749 0.454044
## RACEOther Race              6.796e+00  9.019e-01   7.535 5.21e-14 ***
## age_catyounger              1.077e+00  3.000e-01   3.591 0.000331 ***
## NP                         -2.404e-02  1.050e-01  -0.229 0.818846
## HINCP:RACEAsian             1.169e-05  5.541e-06   2.110 0.034904 *
## HINCP:RACEBlack             8.482e-06  6.029e-06   1.407 0.159540
## HINCP:RACEMultiracial      -4.175e-05  1.144e-05  -3.649 0.000264 ***
## HINCP:RACENative Am.       -1.732e-05  4.448e-05  -0.389 0.696926
## HINCP:RACEOther Race       -9.305e-05  1.066e-05  -8.728  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.7 on 13136 degrees of freedom
## Multiple R-squared:  0.2713,  Adjusted R-squared:  0.2704
## F-statistic: 305.6 on 16 and 13136 DF,  p-value: < 2.2e-16
```

Linear Regression | Best Fit

When looking at educational attainment, the coefficients have changed quite a bit.

One coefficient is statistically significant. In this model, a household with someone with a high school diploma pays, on average, 0.9 percentage points less on rent as a percentage of their household income than someone without a high school diploma.

The other coefficients are not statistically significant. We can assume one reason these results have changed is that we are now controlling for income. Previously, the coefficients could possibly be attributed to the fact that people with higher education have higher incomes.

Now, controlling for income, a household with someone with a Bachelor's Degree or higher pays, on average, 0.6 percentage points more on rent as a percentage of their household income than someone without a high school diploma. But again, the results are not statistically significant.

```
##
## Call:
## lm(formula = GRPIP ~ HINCP + EDU + RACE + age_cat + NP + HINCP:RACE,
##     data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.857  -9.988  -3.553   5.581  87.821
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.996e+01  5.277e-01  75.729  < 2e-16 ***
## HINCP                     -1.217e-04  2.241e-06 -54.306  < 2e-16 ***
## EDUAssoc. Degree          -1.357e+00  6.972e-01  -1.947 0.051615 .
## EDUBach Degree or Higher   5.958e-01  5.117e-01   1.164 0.244333
## EDUHS Diploma             -9.521e-01  4.762e-01  -1.999 0.045575 *
## RACEAsian                  1.632e+00  7.672e-01   2.127 0.033441 *
## RACEBlack                  4.566e-01  6.511e-01   0.701 0.483162
## RACEMultiracial            4.553e+00  1.173e+00   3.882 0.000104 ***
## RACENative Am.            -3.341e+00  4.463e+00  -0.749 0.454044
## RACEOther Race             6.796e+00  9.019e-01   7.535 5.21e-14 ***
## age_catyounger             1.077e+00  3.000e-01   3.591 0.000331 ***
## NP                        -2.404e-02  1.050e-01  -0.229 0.818846
## HINCP:RACEAsian            1.169e-05  5.541e-06   2.110 0.034904 *
## HINCP:RACEBlack            8.482e-06  6.029e-06   1.407 0.159540
## HINCP:RACEMultiracial     -4.175e-05  1.144e-05  -3.649 0.000264 ***
## HINCP:RACENative Am.      -1.732e-05  4.448e-05  -0.389 0.696926
## HINCP:RACEOther Race      -9.305e-05  1.066e-05  -8.728  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.7 on 13136 degrees of freedom
## Multiple R-squared:  0.2713, Adjusted R-squared:  0.2704
## F-statistic: 305.6 on 16 and 13136 DF,  p-value: < 2.2e-16
```

Linear Regression | Best Fit

Let's take a look at Household Income and the interaction term between Household Income and Race.

The negative and statistically significant coefficient for Household Income (HINCP) tells us that as one's income increases, they pay fewer percentage points of their household income on rent. This is controlling for all other variables.

When looking at the interaction term, I am still comparing race to the most common race in my sample set: White alone. Not all coefficients are statistically significant, but let's take a look at a plot to better understand...

```
##
## Call:
## lm(formula = GRPIP ~ HINCP + EDU + RACE + age_cat + NP + HINCP:RACE,
##     data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.857  -9.988  -3.553   5.581  87.821
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.996e+01  5.277e-01  75.729  < 2e-16 ***
## HINCP                     -1.217e-04  2.241e-06 -54.306  < 2e-16 ***
## EDUAssoc. Degree          -1.357e+00  6.972e-01  -1.947 0.051615 .
## EDUBach Degree or Higher   5.958e-01  5.117e-01   1.164 0.244333
## EDUHS Diploma             -9.521e-01  4.762e-01  -1.999 0.045575 *
## RACEAsian                  1.632e+00  7.672e-01   2.127 0.033441 *
## RACEBlack                  4.566e-01  6.511e-01   0.701 0.483162
## RACEMultiracial            4.553e+00  1.173e+00   3.882 0.000104 ***
## RACENative Am.            -3.341e+00  4.463e+00  -0.749 0.454044
## RACEOther Race             6.796e+00  9.019e-01   7.535 5.21e-14 ***
## age_catyounger             1.077e+00  3.000e-01   3.591 0.000331 ***
## NP                        -2.404e-02  1.050e-01  -0.229 0.818846
## HINCP:RACEAsian            1.169e-05  5.541e-06   2.110 0.034904 *
## HINCP:RACEBlack            8.482e-06  6.029e-06   1.407 0.159540
## HINCP:RACEMultiracial     -4.175e-05  1.144e-05  -3.649 0.000264 ***
## HINCP:RACENative Am.      -1.732e-05  4.448e-05  -0.389 0.696926
## HINCP:RACEOther Race      -9.305e-05  1.066e-05  -8.728  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.7 on 13136 degrees of freedom
## Multiple R-squared:  0.2713, Adjusted R-squared:  0.2704
## F-statistic: 305.6 on 16 and 13136 DF,  p-value: < 2.2e-16
```

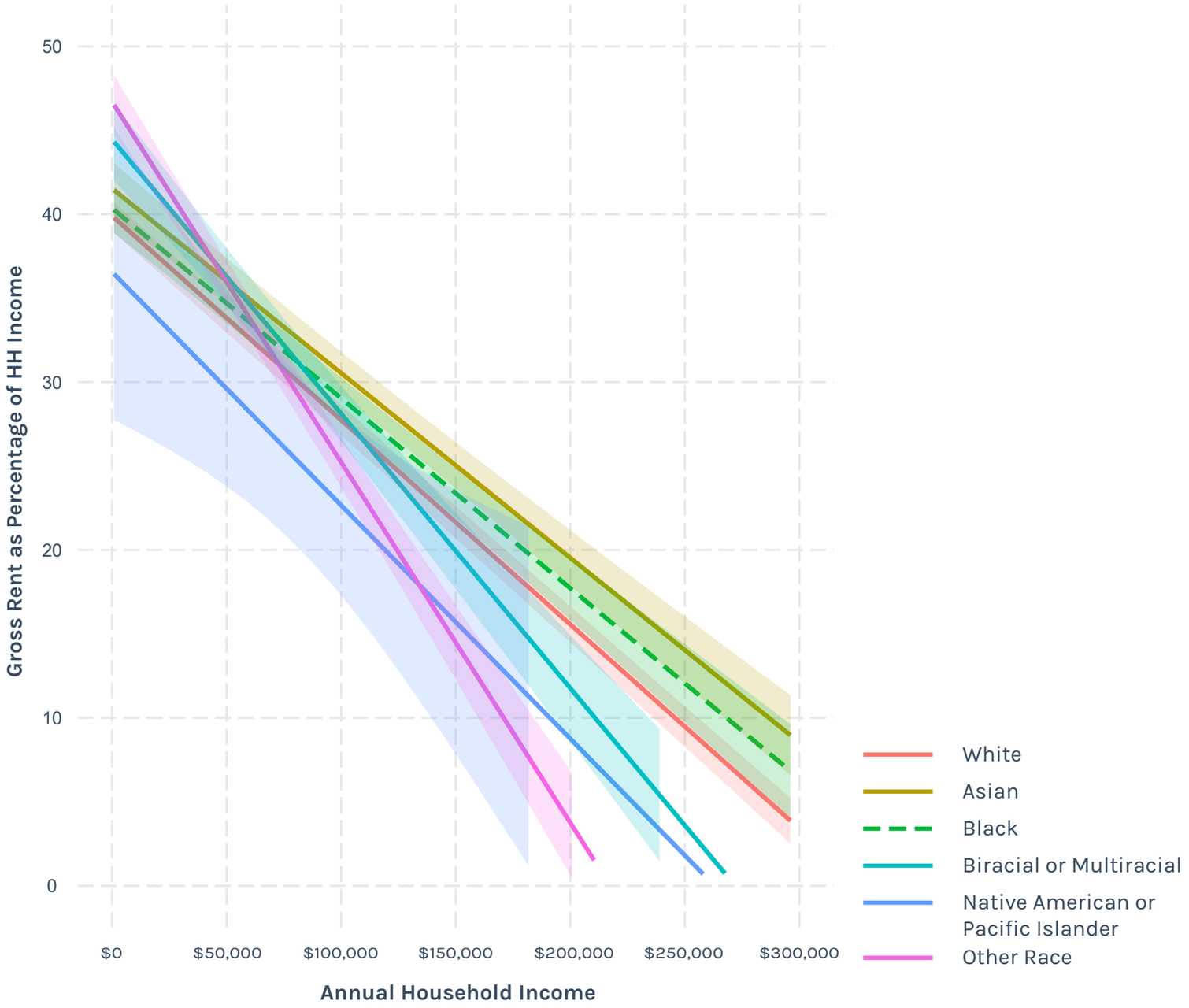
Linear Regression | Best Fit

When looking at middle-income households (renters who make around $50,000), there is not much of a difference in GRPIP between the different racial categories. Everyone pays around 35%-38% of their income on rent. The real differences in the effect of race are seen at very low incomes and very high incomes.

Let's take Asian renters as an example. The positive and statistically significant coefficient for the interaction between this race and household income tells us that income has slightly more of an effect on GRPIP for Asian renters than it does for White renters. As we see here, the slopes representing White renters and Asian renters start off around the same point, but as we move up to a higher household income ($300,000), there is a bigger difference between the GRPIP of White and Asian renters.

The same is true when comparing White renters and Black renters.

Oppositely, there is a negative and statistically significant coefficient for the interaction between household income and multiracial as well as some other race.



Linear Regression | Best Fit

Finally, we should quickly take a look at our new variable. If you are 30 years old or younger ,you pay 1 percentage point more on rent than if you are over the age of 30. This coefficient is statistically significant.

```
##
## Call:
## lm(formula = GRPIP ~ HINCP + EDU + RACE + age_cat + NP + HINCP:RACE,
##     data = person_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.857  -9.988  -3.553   5.581  87.821
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.996e+01  5.277e-01  75.729  < 2e-16 ***
## HINCP                      -1.217e-04  2.241e-06 -54.306  < 2e-16 ***
## EDUAssoc. Degree           -1.357e+00  6.972e-01  -1.947 0.051615 .
## EDUBach Degree or Higher    5.958e-01  5.117e-01   1.164 0.244333
## EDUHS Diploma              -9.521e-01  4.762e-01  -1.999 0.045575 *
## RACEAsian                   1.632e+00  7.672e-01   2.127 0.033441 *
## RACEBlack                   4.566e-01  6.511e-01   0.701 0.483162
## RACEMultiracial             4.553e+00  1.173e+00   3.882 0.000104 ***
## RACENative Am.             -3.341e+00  4.463e+00  -0.749 0.454044
## RACEOther Race              6.796e+00  9.019e-01   7.535 5.21e-14 ***
## age_catyounger              1.077e+00  3.000e-01   3.591 0.000331 ***
## NP                         -2.404e-02  1.050e-01  -0.229 0.818846
## HINCP:RACEAsian             1.169e-05  5.541e-06   2.110 0.034904 *
## HINCP:RACEBlack             8.482e-06  6.029e-06   1.407 0.159540
## HINCP:RACEMultiracial      -4.175e-05  1.144e-05  -3.649 0.000264 ***
## HINCP:RACENative Am.       -1.732e-05  4.448e-05  -0.389 0.696926
## HINCP:RACEOther Race       -9.305e-05  1.066e-05  -8.728  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.7 on 13136 degrees of freedom
## Multiple R-squared:  0.2713, Adjusted R-squared:  0.2704
## F-statistic: 305.6 on 16 and 13136 DF,  p-value: < 2.2e-16
```

Research Question and Hypothesis

Some of my hypotheses were confirmed, but maybe not to the extent I was expecting. Some were not. In general, we saw different results throughout the various statistical tests. Let’s take a look at each hypothesis:

Discussion

Research Question and Hypothesis

Do people in minority racial groups (people who do not identify as white) pay a higher gross rent as a percentage of household income?

It depends on what other variables you are controlling for. In the final linear regression model, we see that non-majority racial groups, other than Native American or Pacific Islander, pay more on rent than white renters. However only the ‘Asian’, ‘biracial or multiracial’, and ‘some other race’ coefficients are statistically significant, so we can conclude that those groups pay more on rent when controlling for all of the other variables listed.

Do people with a lower level of education have a higher gross rent as a percentage of household income.

Again, it of course depends on what other variables you are controlling for. I thought the two-sample t-test was interesting in showing that, on average, people without a Bachelor’s Degree are more rent-burdened. We also saw, in the first linear regression model, that people with higher levels of education pay fewer percentage points on rent as a percentage of household income as compared to people with no high school diploma. But in the final linear regression, this was not the case.

Do lower-income households have higher gross rents as a percentage of household income?

Yes, generally. We saw in the linear regression that as one’s income increases, they pay fewer percentage points of their household income on rent.

Research Question and Hypothesis

There are a few limitations to my study. First is the issue of **person-level data vs. household-level data.** Would there be differences in my results if I were to incorporate, for example, all-Black households or all-Asian households? Or, a household where everyone has a Master's degree or higher? By examining these variables at a person level, I may be missing out on some nuances or greater differences between household demographics.

My **categorization** of educational level and race is **bias** and perhaps too overarching. These groupings could be refined, rethought, and tested in various ways to understand a more complete picture.

Finally, the next step would be to **aggregate these down to the census tract level** to understand which parts of the state are more-rent burdened or where rent burdened people live.

Conclusion

Research Question and Hypothesis

An analysis of rent burden at the state level is important to understand what economic challenges our community faces. Massachusetts has one of the highest costs of living and the highest average housing costs. Understanding which demographics are more likely to be rent-burdened could allow us to address these populations directly and develop policies or support systems to ensure no one is spending more than 30% of their income on rent.

Understanding how income and educational obtainment affects rent burden is crucial. Generally, Massachusetts of highly-education individuals. It is also a place of great disparity. We need to understand the financial implication higher education may have on individuals and households (i.e. student debt) as well as the inequalities for those without higher education. Which are the groups that are being forgotten when it comes to addressing financial security and rent burden?

Finally, we know that racial inequality in housing is a systemic issue. There are many reasons why minority racial groups may pay higher percentages of their income on rent, and these factors are entrenched in our white supremacist society. We must work to solve this. By seeing which demographics are rent-burdened, we can address this at a regional, state and national level.