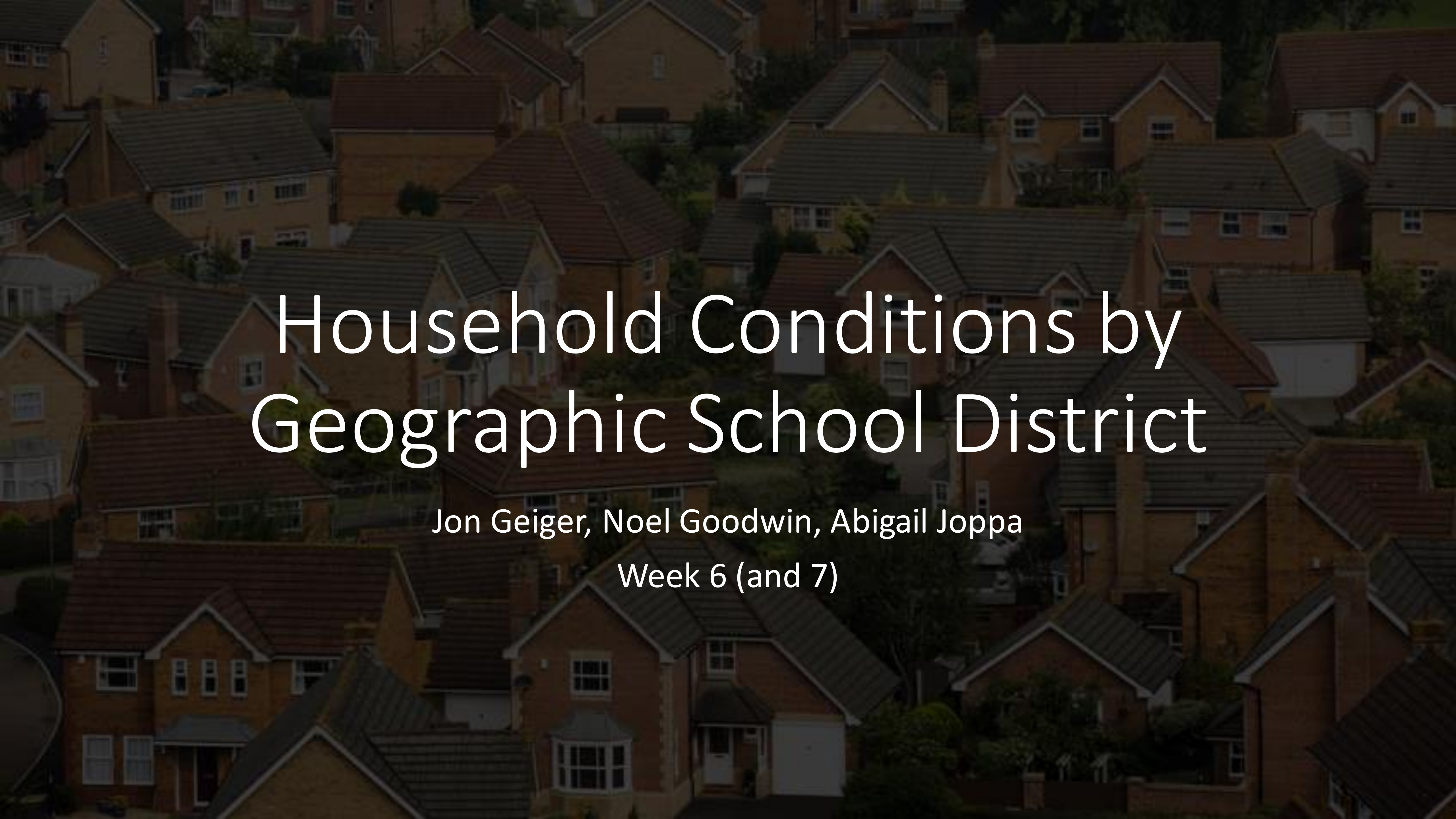


Household Conditions by Geographic School District

Jon Geiger, Noel Goodwin, Abigail Joppa

Week 6 (and 7)

Last Week's Research Question

One of the only measures being used to predict graduation rates are GPA, tests, disciplinary issues, attendance etc. using a tracking system called EWS (early warning system).

However, we know that household conditions significantly impact these indicators. This leads us to the question:

What household conditions are the biggest indicators for graduation rates across school districts?

How does this differ across:

- ➢Racial and ethnic groups?
- ➢Regions?
- ➢Household income groups?

Last Week's Next Steps…

Join schools to the data containing graduation rate predictors.

Creating a district wide assessment measure for the indicators (weighted by the number of students in each of the schools)

Build regression model to predict graduation rates using HH conditions

Research Question Difficulties

- Difficult to "group by" racial/ethnic groups
- Self-defining regions: doable
- Household income group data
  - Proxy with school district funding?
  - Look not at income groups, but ranked groups of amount of funding

Joining Schools Data Containing Grad Rate Predictors

Traditional Indicators: Attendance, Grades, Assessments, Discipline

Education Data Package in R provides the following datasets by school:

- Discipline Instances (2015, 2017)
  - Missing Data
- Chronic Absenteeism (2013, 2015)
  - Huge download
- Assessments (2009-2018)
  - Math and Reading Test Scores
  - Data available for grades 3-9 or "99"

Is this data usable for a regression model?

District-Wide Assessment Indicators

Issues:

- Missing Data/Years
- Fundamentally different model (Classification vs. Regression)

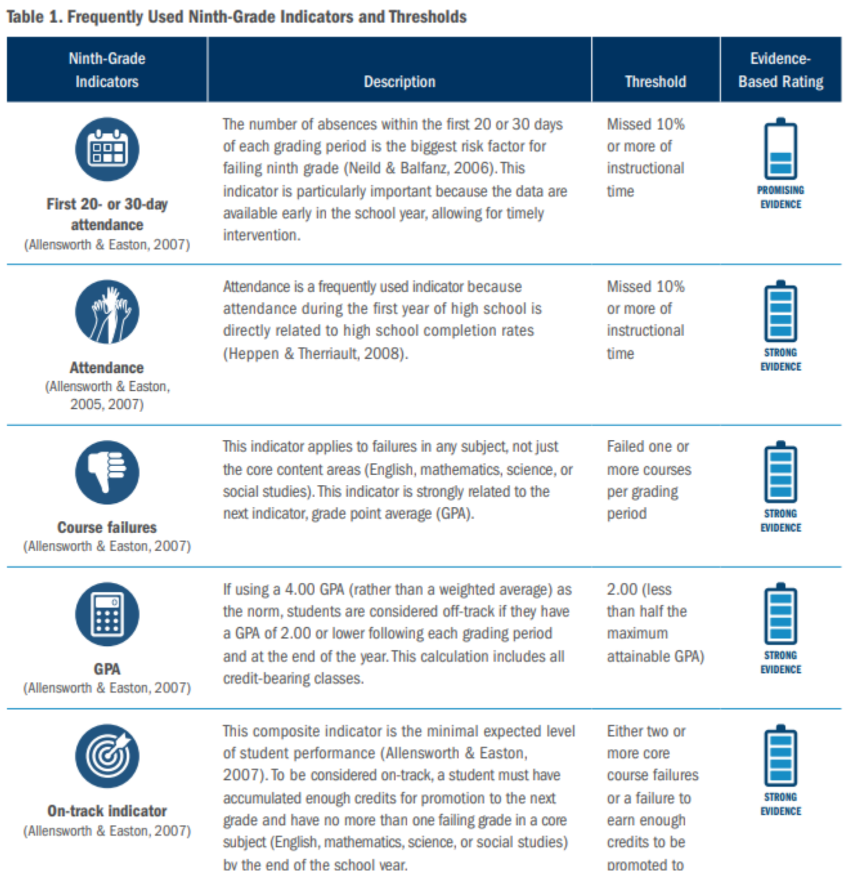


Table 1. Frequently Used Ninth-Grade Indicators and Thresholds

Ninth-Grade Indicators	Description	Threshold	Evidence-Based Rating
First 20- or 30-day attendance (Allensworth & Easton, 2007)	The number of absences within the first 20 or 30 days of each grading period is the biggest risk factor for failing ninth grade (Neild & Balfanz, 2006). This indicator is particularly important because the data are available early in the school year, allowing for timely intervention.	Missed 10% or more of instructional time	PROMISING EVIDENCE
Attendance (Allensworth & Easton, 2005, 2007)	Attendance is a frequently used indicator because attendance during the first year of high school is directly related to high school completion rates (Heppen & Therriault, 2008).	Missed 10% or more of instructional time	STRONG EVIDENCE
Course failures (Allensworth & Easton, 2007)	This indicator applies to failures in any subject, not just the core content areas (English, mathematics, science, or social studies). This indicator is strongly related to the next indicator, grade point average (GPA).	Failed one or more courses per grading period	STRONG EVIDENCE
GPA (Allensworth & Easton, 2007)	If using a 4.00 GPA (rather than a weighted average) as the norm, students are considered off-track if they have a GPA of 2.00 or lower following each grading period and at the end of the year. This calculation includes all credit-bearing classes.	2.00 (less than half the maximum attainable GPA)	STRONG EVIDENCE
On-track indicator (Allensworth & Easton, 2007)	This composite indicator is the minimal expected level of student performance (Allensworth & Easton, 2007). To be considered on-track, a student must have accumulated enough credits for promotion to the next grade and have no more than one failing grade in a core subject (English, mathematics, science, or social studies) by the end of the school year.	Either two or more core course failures or a failure to earn enough credits to be promoted to	STRONG EVIDENCE

Regression!

Linear Regression:

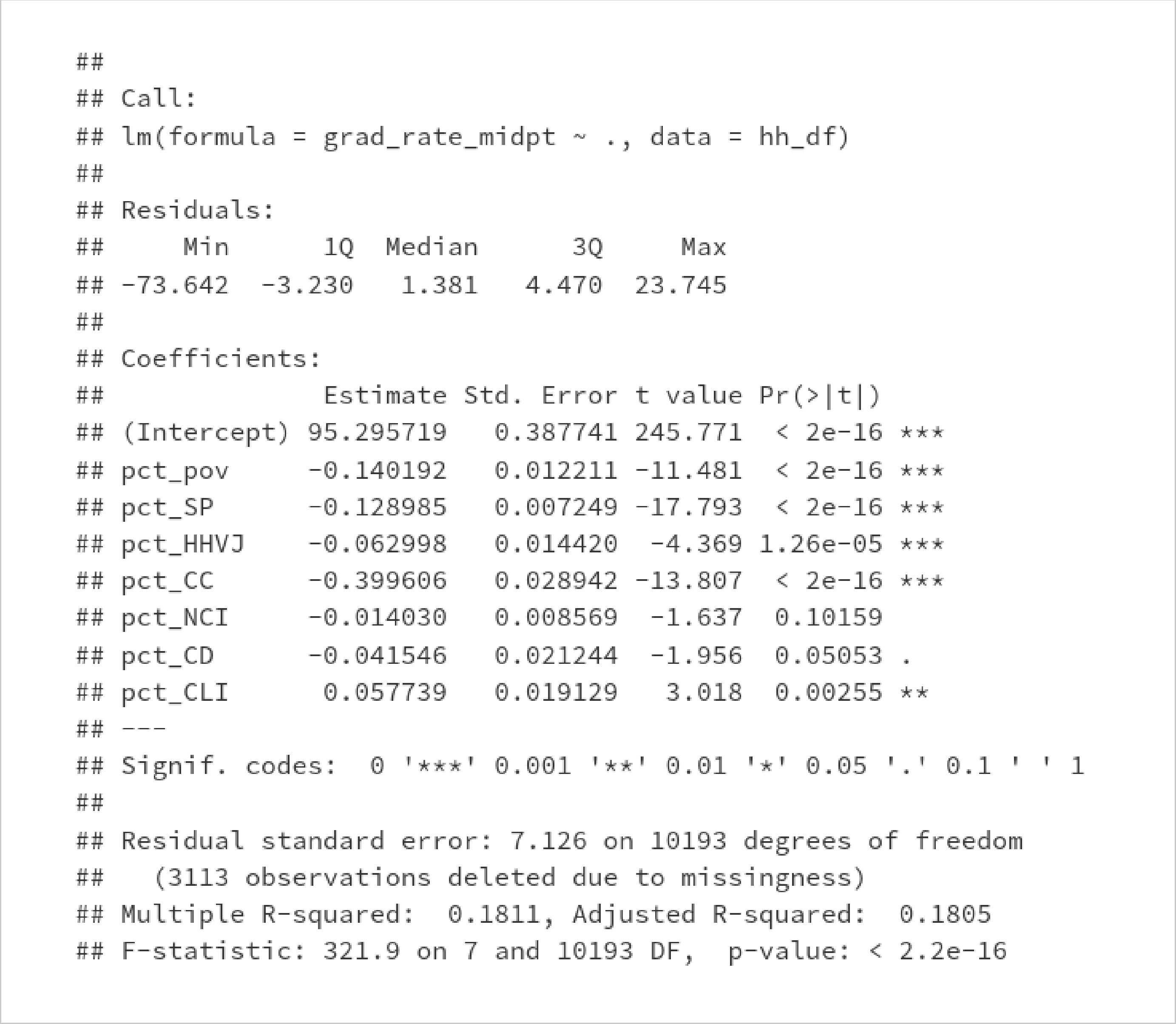
1. Household Condition Data
2. Race Data
   1. %'s of Hispanic/Latino, White, Black, Native American, Asian, and Pacific Islander
3. Combined

ANOVA to test whether adding race data significantly improves the RSS

Regression!

Graduation Rate based on:

Household Conditions

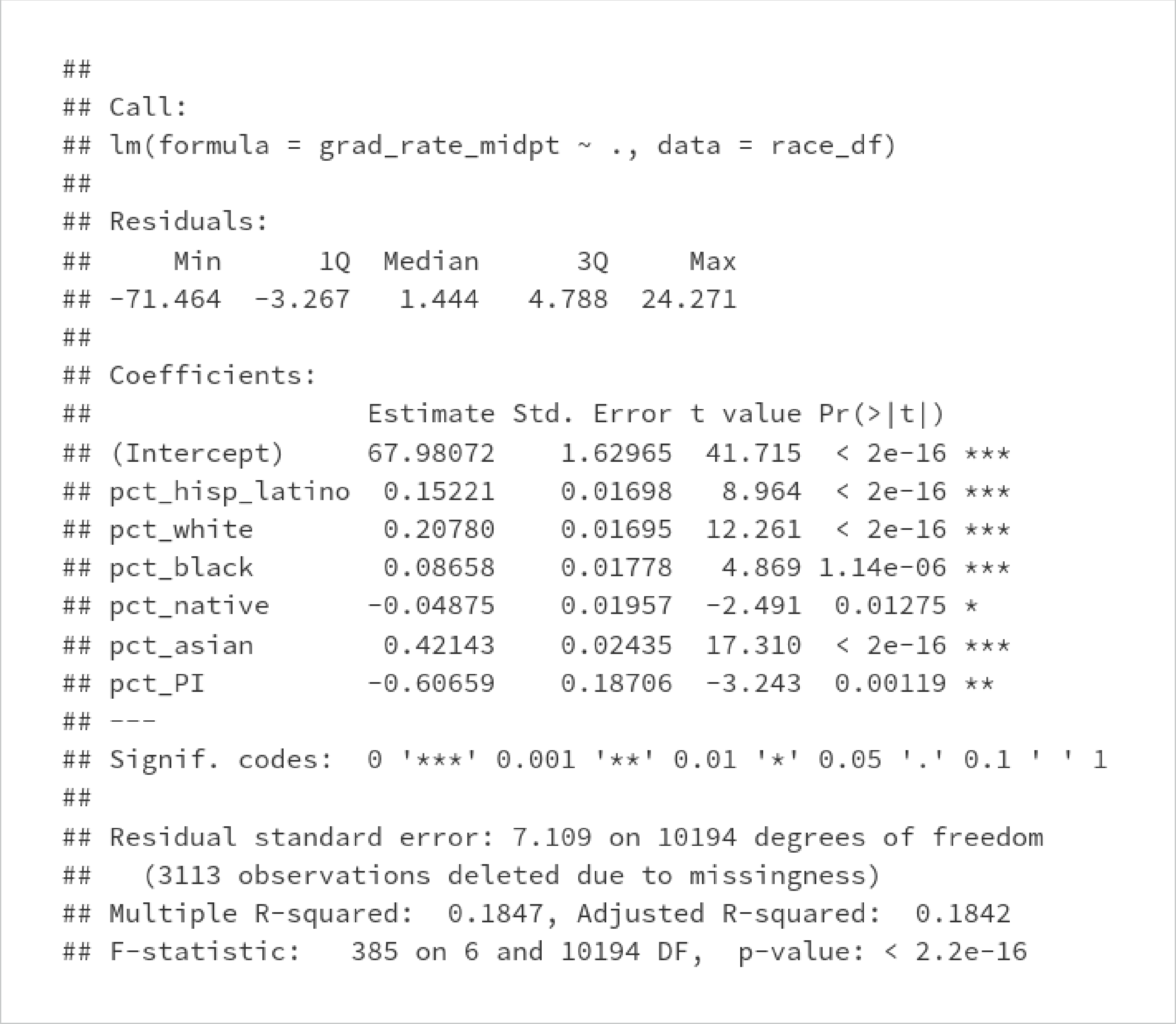


```
##
## Call:
## lm(formula = grad_rate_midpt ~ ., data = hh_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.642  -3.230   1.381   4.470  23.745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 95.295719   0.387741 245.771  < 2e-16 ***
## pct_pov     -0.140192   0.012211 -11.481  < 2e-16 ***
## pct_SP      -0.128985   0.007249 -17.793  < 2e-16 ***
## pct_HHVJ    -0.062998   0.014420  -4.369 1.26e-05 ***
## pct_CC      -0.399606   0.028942 -13.807  < 2e-16 ***
## pct_NCI     -0.014030   0.008569  -1.637  0.10159
## pct_CD      -0.041546   0.021244  -1.956  0.05053 .
## pct_CLI      0.057739   0.019129   3.018  0.00255 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.126 on 10193 degrees of freedom
##   (3113 observations deleted due to missingness)
## Multiple R-squared:  0.1811, Adjusted R-squared:  0.1805
## F-statistic: 321.9 on 7 and 10193 DF,  p-value: < 2.2e-16
```

Regression!

Graduation Rate based on:

Racial makeup

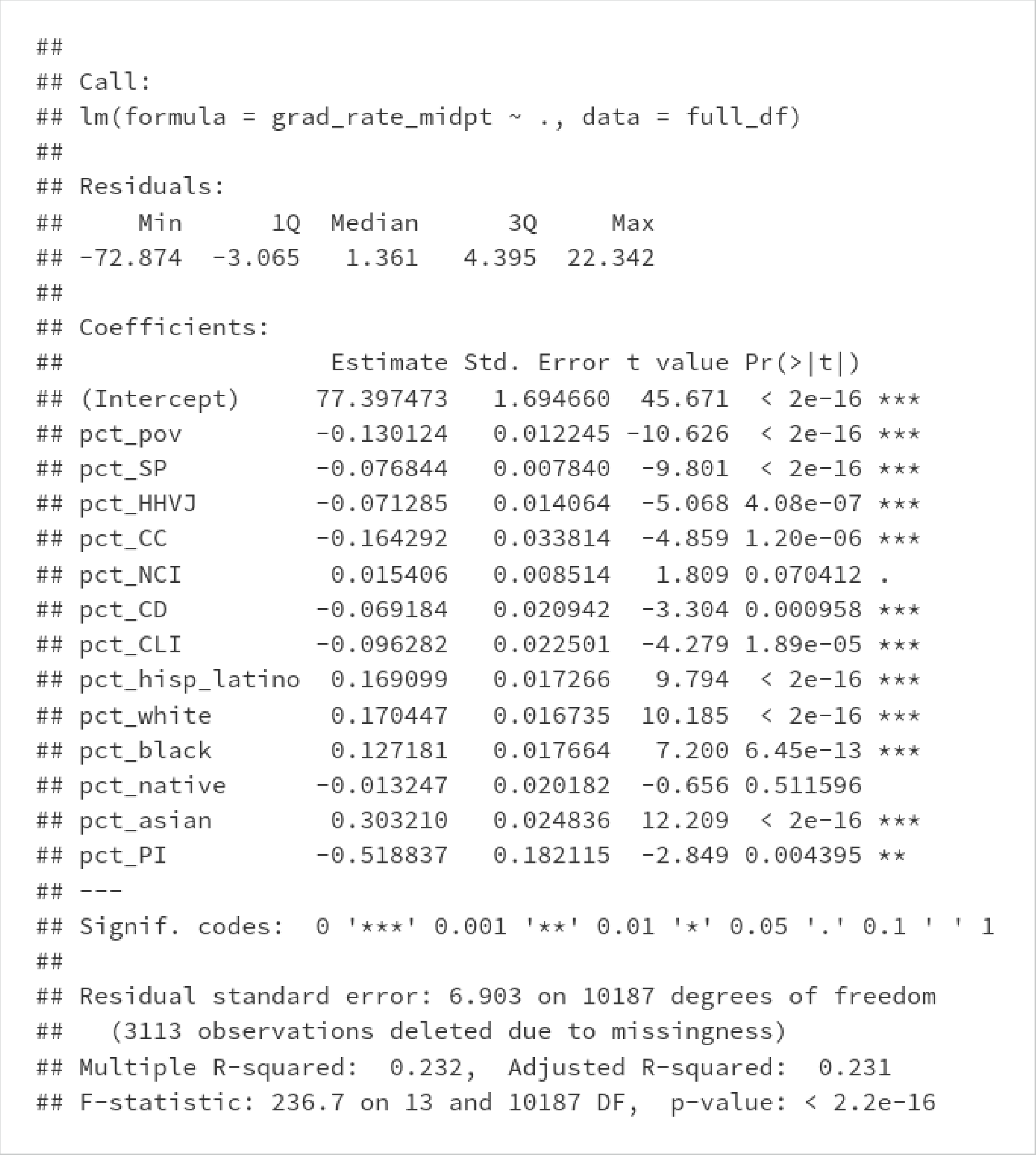


```
##
## Call:
## lm(formula = grad_rate_midpt ~ ., data = race_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -71.464  -3.267   1.444   4.788  24.271
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     67.98072    1.62965  41.715  < 2e-16 ***
## pct_hisp_latino  0.15221    0.01698   8.964  < 2e-16 ***
## pct_white        0.20780    0.01695  12.261  < 2e-16 ***
## pct_black        0.08658    0.01778   4.869 1.14e-06 ***
## pct_native      -0.04875    0.01957  -2.491  0.01275 *
## pct_asian        0.42143    0.02435  17.310  < 2e-16 ***
## pct_PI          -0.60659    0.18706  -3.243  0.00119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.109 on 10194 degrees of freedom
##   (3113 observations deleted due to missingness)
## Multiple R-squared:  0.1847, Adjusted R-squared:  0.1842
## F-statistic:   385 on 6 and 10194 DF,  p-value: < 2.2e-16
```

Regression!

Graduation Rate based on:

Household Conditions + Racial Makeup

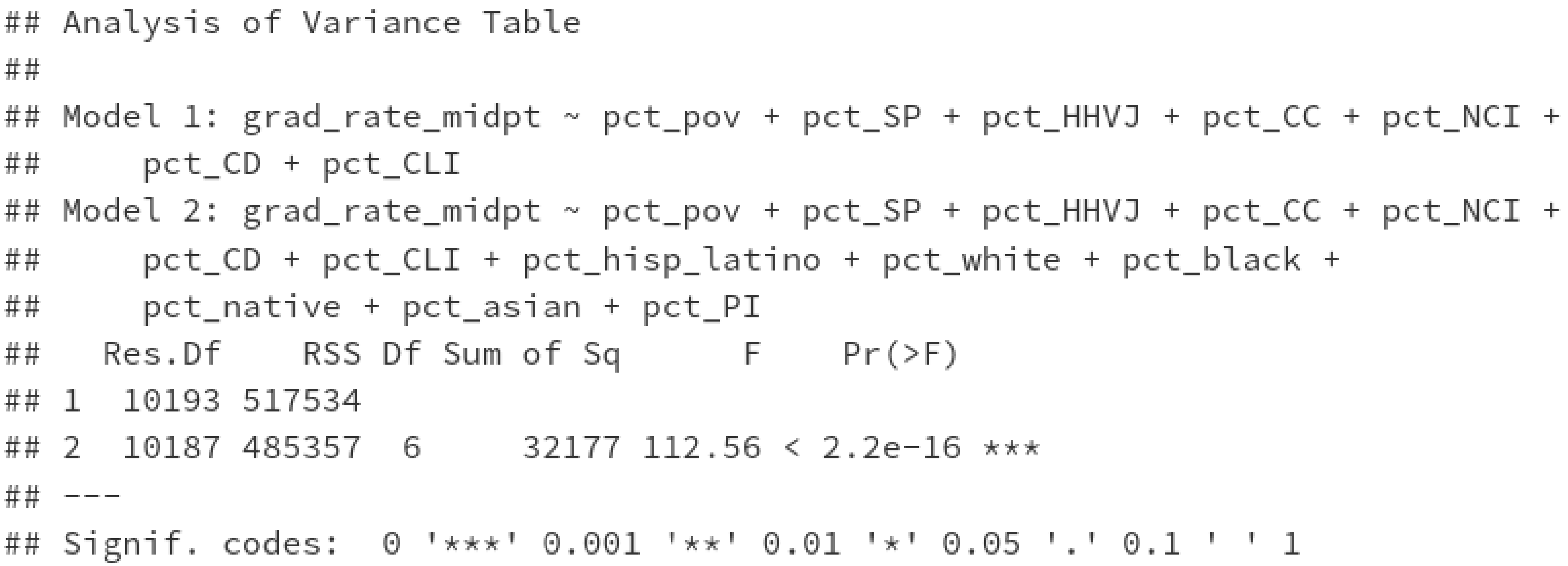


```
##
## Call:
## lm(formula = grad_rate_midpt ~ ., data = full_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -72.874  -3.065   1.361   4.395  22.342
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     77.397473   1.694660  45.671  < 2e-16 ***
## pct_pov         -0.130124   0.012245 -10.626  < 2e-16 ***
## pct_SP          -0.076844   0.007840  -9.801  < 2e-16 ***
## pct_HHVJ        -0.071285   0.014064  -5.068 4.08e-07 ***
## pct_CC          -0.164292   0.033814  -4.859 1.20e-06 ***
## pct_NCI          0.015406   0.008514   1.809 0.070412 .
## pct_CD          -0.069184   0.020942  -3.304 0.000958 ***
## pct_CLI         -0.096282   0.022501  -4.279 1.89e-05 ***
## pct_hisp_latino  0.169099   0.017266   9.794  < 2e-16 ***
## pct_white        0.170447   0.016735  10.185  < 2e-16 ***
## pct_black        0.127181   0.017664   7.200 6.45e-13 ***
## pct_native      -0.013247   0.020182  -0.656 0.511596
## pct_asian        0.303210   0.024836  12.209  < 2e-16 ***
## pct_PI          -0.518837   0.182115  -2.849 0.004395 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.903 on 10187 degrees of freedom
##   (3113 observations deleted due to missingness)
## Multiple R-squared:  0.232,  Adjusted R-squared:  0.231
## F-statistic: 236.7 on 13 and 10187 DF,  p-value: < 2.2e-16
```

Regression!

Comparison of:

- Household Conditions model and
- Combined Model

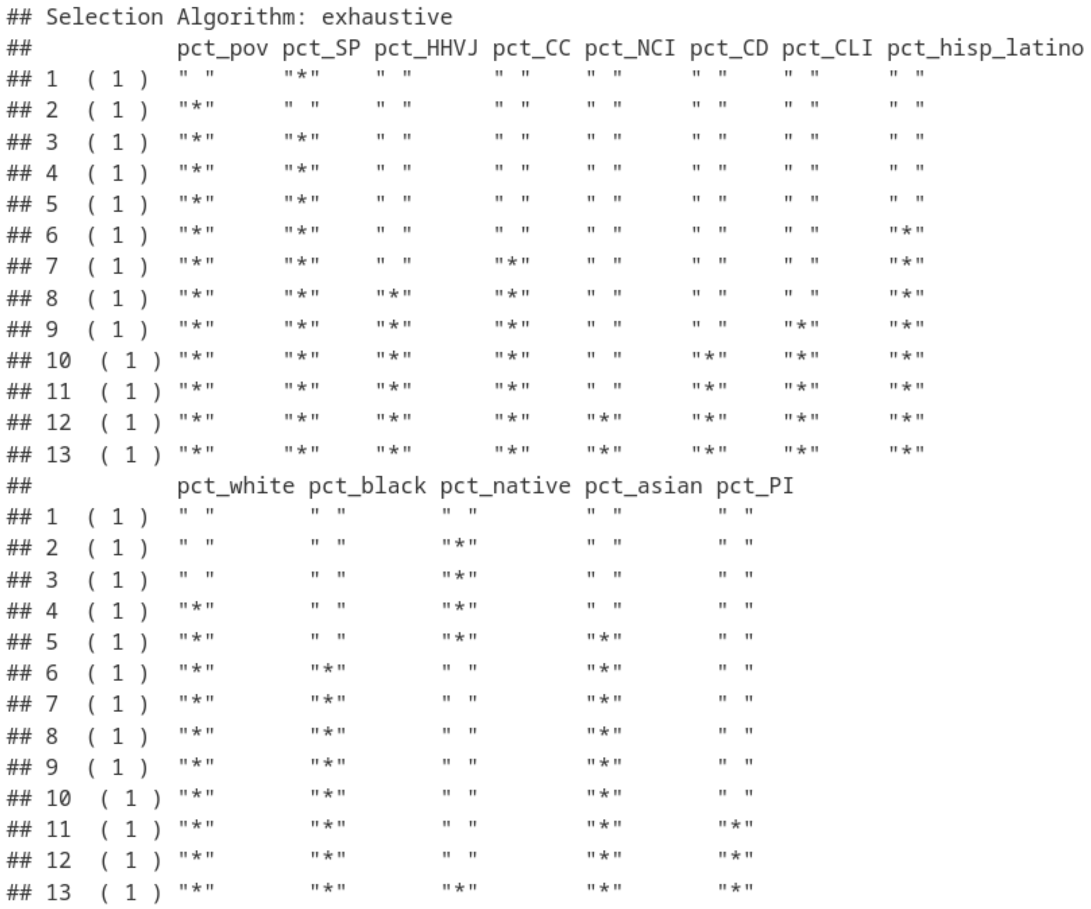


```
## Analysis of Variance Table
##
## Model 1: grad_rate_midpt ~ pct_pov + pct_SP + pct_HHVJ + pct_CC + pct_NCI +
##     pct_CD + pct_CLI
## Model 2: grad_rate_midpt ~ pct_pov + pct_SP + pct_HHVJ + pct_CC + pct_NCI +
##     pct_CD + pct_CLI + pct_hisp_latino + pct_white + pct_black +
##     pct_native + pct_asian + pct_PI
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  10193 517534
## 2  10187 485357  6     32177 112.56 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

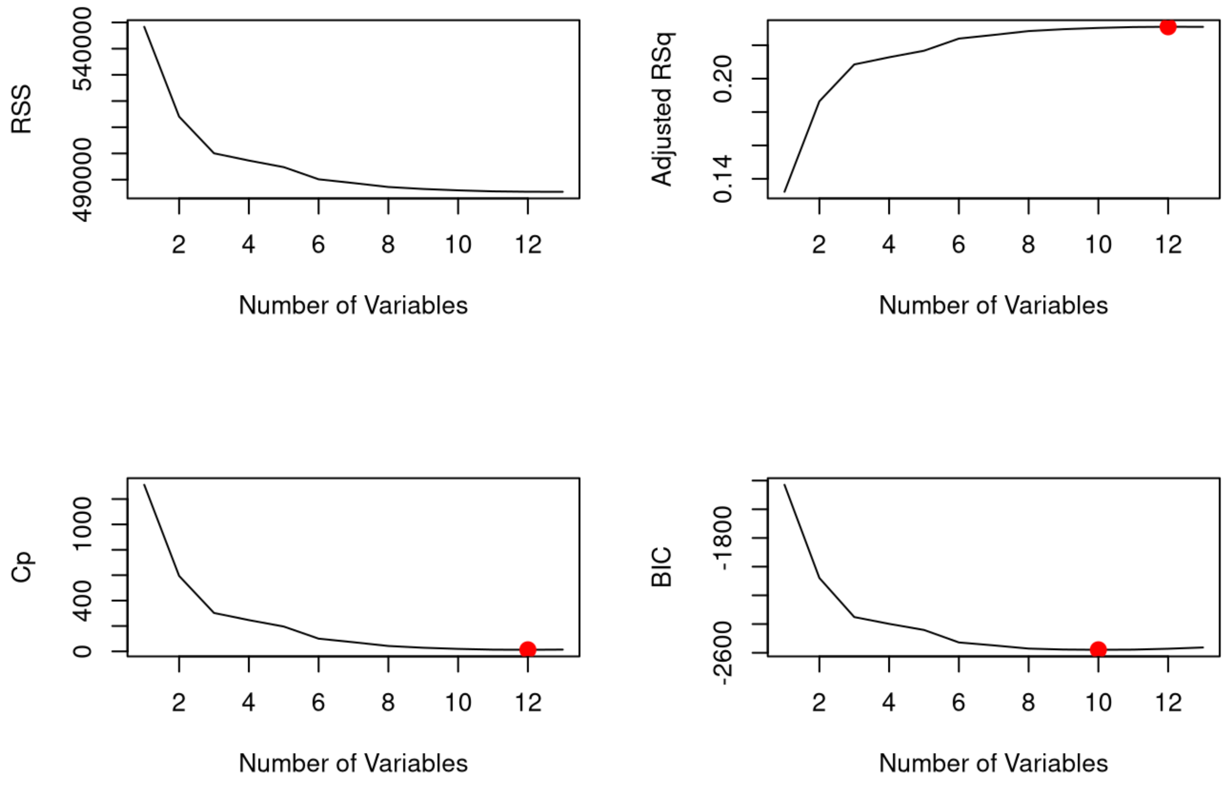
Regression!

Variable Selection:

Best Subset Selection



```
## Selection Algorithm: exhaustive
##           pct_pov pct_SP pct_HHVJ pct_CC pct_NCI pct_CD pct_CLI pct_hisp_latino
## 1  ( 1 )  " "     "*"    " "      " "    " "     " "    " "     " "
## 2  ( 1 )  "*"     " "    " "      " "    " "     " "    " "     " "
## 3  ( 1 )  "*"     "*"    " "      " "    " "     " "    " "     " "
## 4  ( 1 )  "*"     "*"    " "      " "    " "     " "    " "     " "
## 5  ( 1 )  "*"     "*"    " "      " "    " "     " "    " "     " "
## 6  ( 1 )  "*"     "*"    " "      " "    " "     " "    " "     "*"
## 7  ( 1 )  "*"     "*"    " "      "*"    " "     " "    " "     "*"
## 8  ( 1 )  "*"     "*"    "*"      "*"    " "     " "    " "     "*"
## 9  ( 1 )  "*"     "*"    "*"      "*"    " "     " "    "*"     "*"
## 10  ( 1 ) "*"     "*"    "*"      "*"    " "     "*"    "*"     "*"
## 11  ( 1 ) "*"     "*"    "*"      "*"    " "     "*"    "*"     "*"
## 12  ( 1 ) "*"     "*"    "*"      "*"    "*"     "*"    "*"     "*"
## 13  ( 1 ) "*"     "*"    "*"      "*"    "*"     "*"    "*"     "*"
##           pct_white pct_black pct_native pct_asian pct_PI
## 1  ( 1 )  " "       " "       " "        " "       " "
## 2  ( 1 )  " "       " "       "*"        " "       " "
## 3  ( 1 )  " "       " "       "*"        " "       " "
## 4  ( 1 )  "*"       " "       "*"        " "       " "
## 5  ( 1 )  "*"       " "       "*"        "*"       " "
## 6  ( 1 )  "*"       "*"       " "        "*"       " "
## 7  ( 1 )  "*"       "*"       " "        "*"       " "
## 8  ( 1 )  "*"       "*"       " "        "*"       " "
## 9  ( 1 )  "*"       "*"       " "        "*"       " "
## 10  ( 1 ) "*"       "*"       " "        "*"       " "
## 11  ( 1 ) "*"       "*"       " "        "*"       "*"
## 12  ( 1 ) "*"       "*"       " "        "*"       "*"
## 13  ( 1 ) "*"       "*"       "*"        "*"       "*"
```



New Research Question

We know that correlations between graduation rate and household/demographic conditions vary by region.

What household conditions are the biggest indicators for graduation rates across school districts?

How does this differ across:

- ➢ Regions?
- ➢ Tiers of school district funding?

Does district-wide assessment data provide a significant improvement to a regression model?

Next Steps...

Visualize predictor distributions for regression

Improve regression model predicting graduation rates using HH conditions and race

Continue to look for prior research that looks at graduation rates and indicators