Unpaid Lunch Debt in Durham, NC

Julia Donheiser

11/2/2018

Introduction

When families can't afford to pay for student lunches, school districts foot the bill. But with major cuts to educational funding in North Carolina—where some schools don't even have enough funds to pay for students' textbooks—this means school districts can wrack up tens of thousands of dollars in debt. In Durham, students with five or more unpaid lunches only receive a juice and a sandwich instead of a hot lunch. This lends its way to "lunch shaming", where students who can't afford pay skip the meal altogether to avoid the embarrassment of eating a cold lunch. This is a major issue, since student performance in school is directly tied to access to quality food.

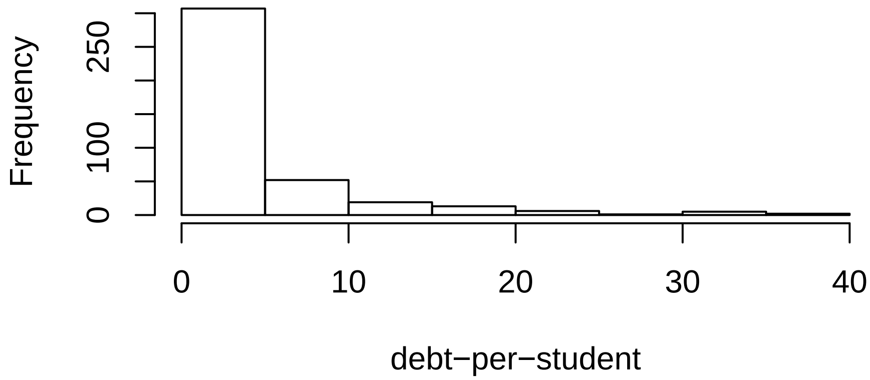
Data Sources

- End-of-Year unpaid meal data from James Keaten, director of child nutrition services at DPS.
- All free/reduced price lunch data was obtained from ncpublicschools.org
- 2010-11 through 2015-16 demographic data was obtained from the NCES ELSI table generator, code 91803
- 2017-18 ADM data from ncpublicschools.org's Average Daily Membership and Membership Last Day by School
- 2016-17, 2017-18 demographic data from Durham Public Schools

Exploratory Data Analysis

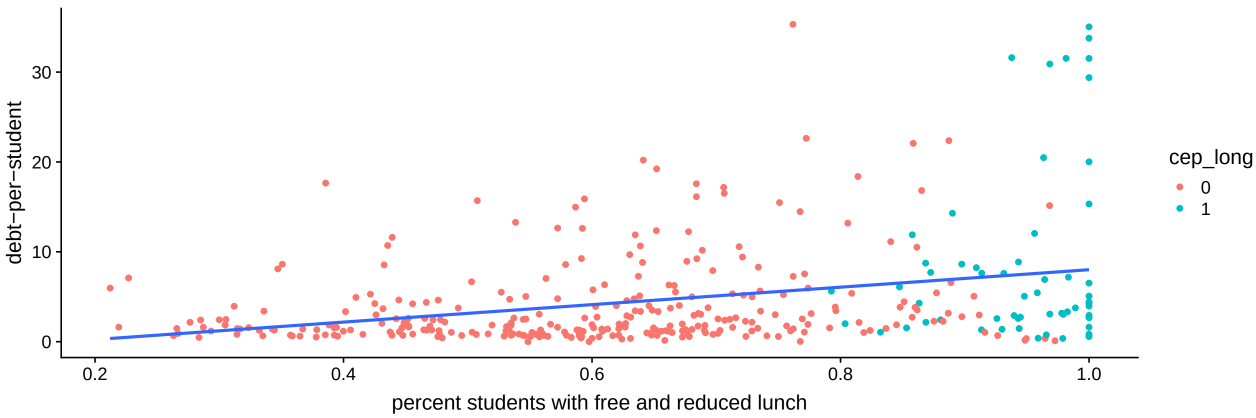
```
## # A tibble: 8 x 4
##   year    total_debt mean_debt_per_student fullprice_lunches
##   <chr>        <dbl>                 <dbl>             <dbl>
## 1 2010-11    204692.                  7.34            70584.
## 2 2011-12    111567.                  4.53            38472.
## 3 2012-13    117526.                  4.54            40526.
## 4 2013-14    108231.                  3.76            37321.
## 5 2014-15     85093.                  2.51            29342.
## 6 2015-16     78428.                  2.25            27044.
## 7 2016-17    127940.                  3.38            44117.
## 8 2017-18    209022.                  5.10            72076.
```

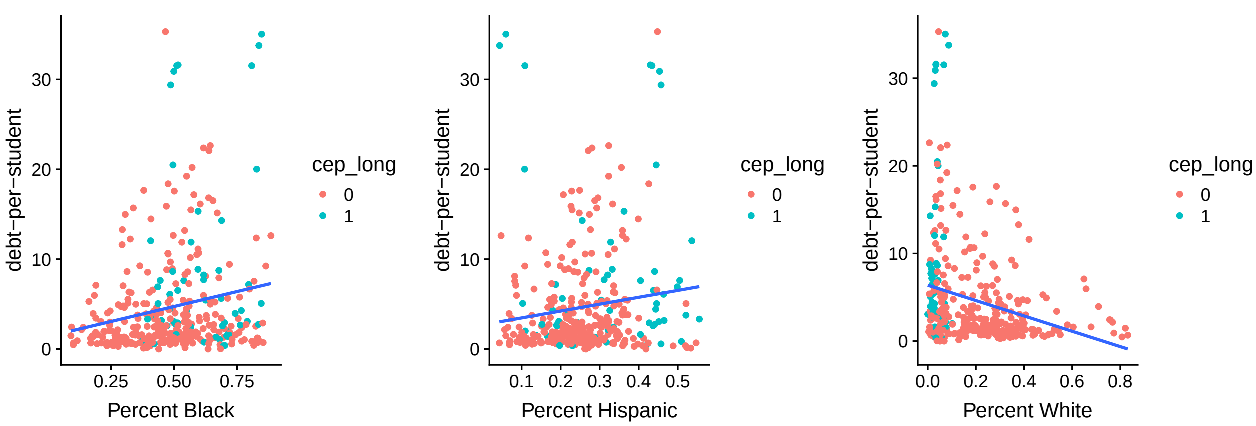
At the end of the 2017-18 academic year, DPS had over $209,000 in school lunch debt. That's over 72,000 unpaid lunches, with an average of $5.10 of debt per student. It's also the most debt the school district has seen in the past eight years.

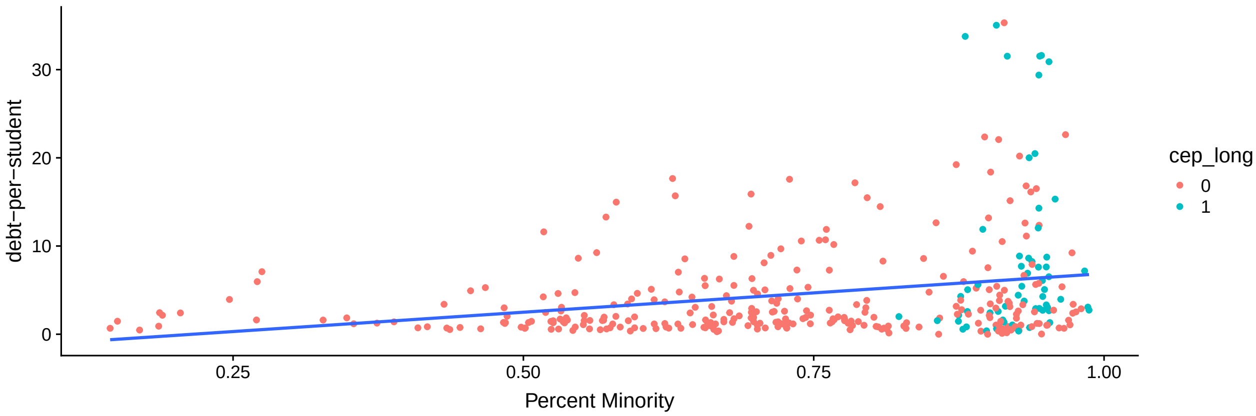


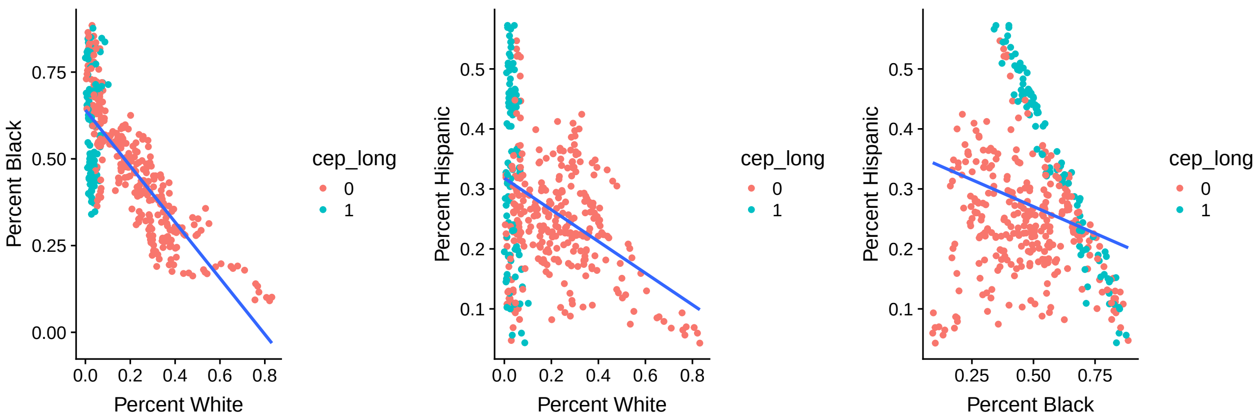
Most schools have less than $5 of lunch debt per student. In Durham, a full-priced lunch costs $2.90, and a reduced-price lunch cost $0.40, according to the Durham Public Schools website. That's about two unpaid full-priced lunches per student, or just over 12 unpaid reduced-price lunches per student. For the rest of my EDA, I'll delve into which schools have more debt and whether we can find systematic issues. I'll also be looking at schools that are part of the Community Eligibility Provision, which means all students receive free lunch.

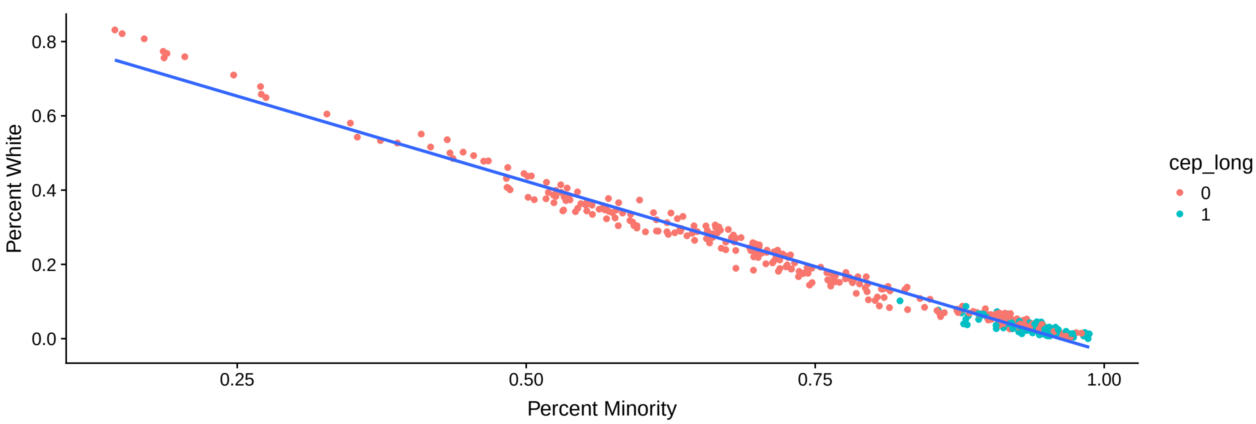
Debt and Demographics





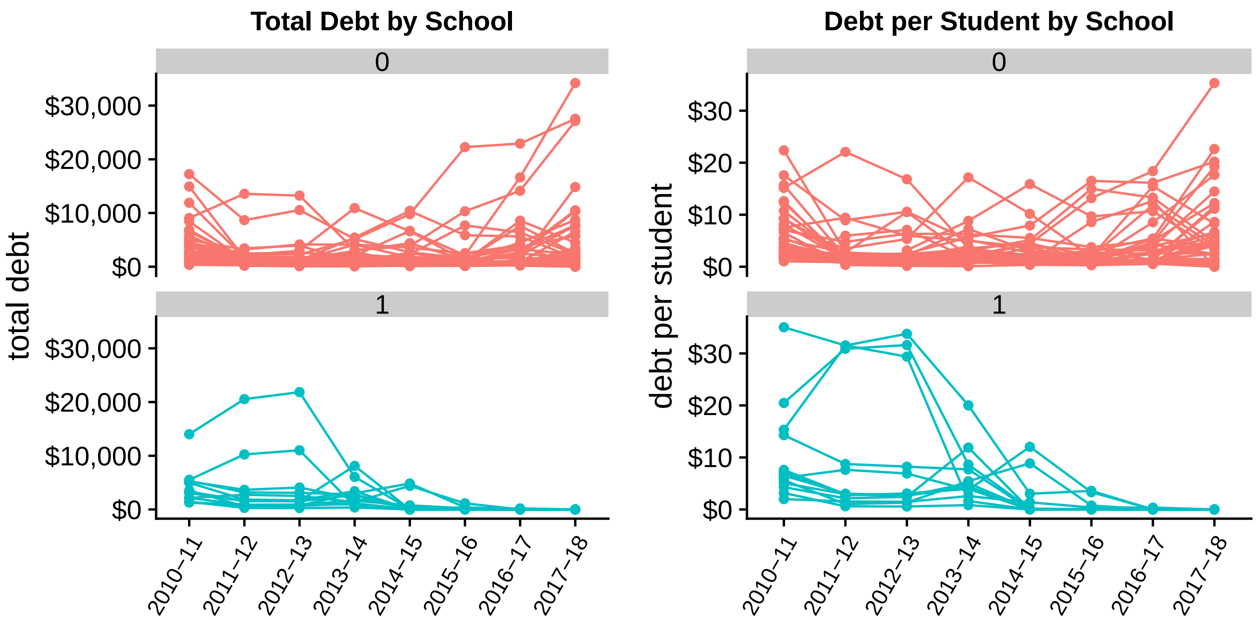






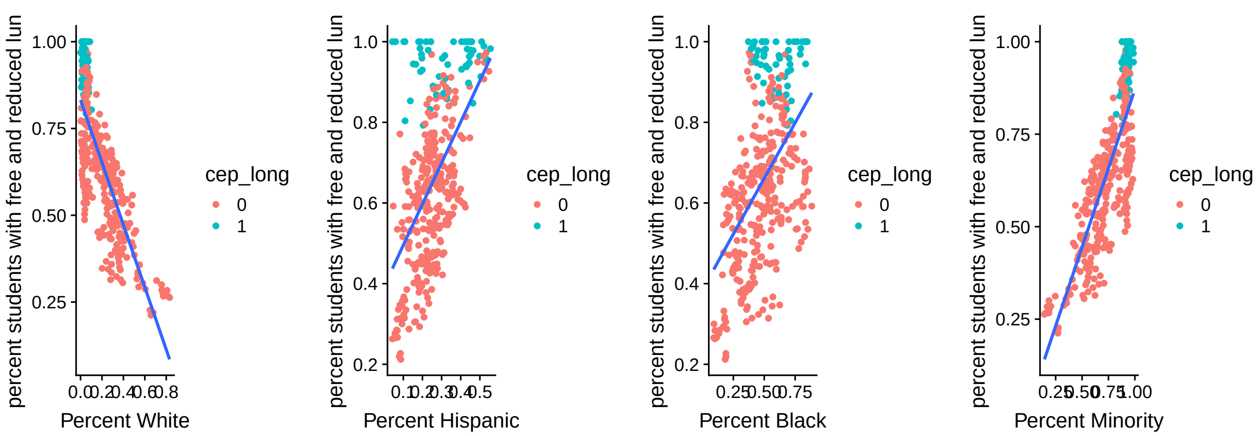
There is a weak positive correlation between the percentage of students who have free and reduced price lunch and the debt-per-student at each school. It looks like schools with CEP status generally have a higher percentage of students on free/reduced lunch. As for demographics, there is also a weak positive correlation between the percentage of black students and debt-per-student. Conversely, there is a weak negative correlation between the percent of white students and debt-per-student. There doesn't appear to be a much of a relationship between the percentage of hispanic students and debt-per-student. That being said, if we look at the total percentage of minority students — the percentage of black *and* hispanic students — there is a positive correlation with debt-per-student. Schools that CEP status are also starkly segregated, with very few white students and primarily black and hispanic students.

Debt over time



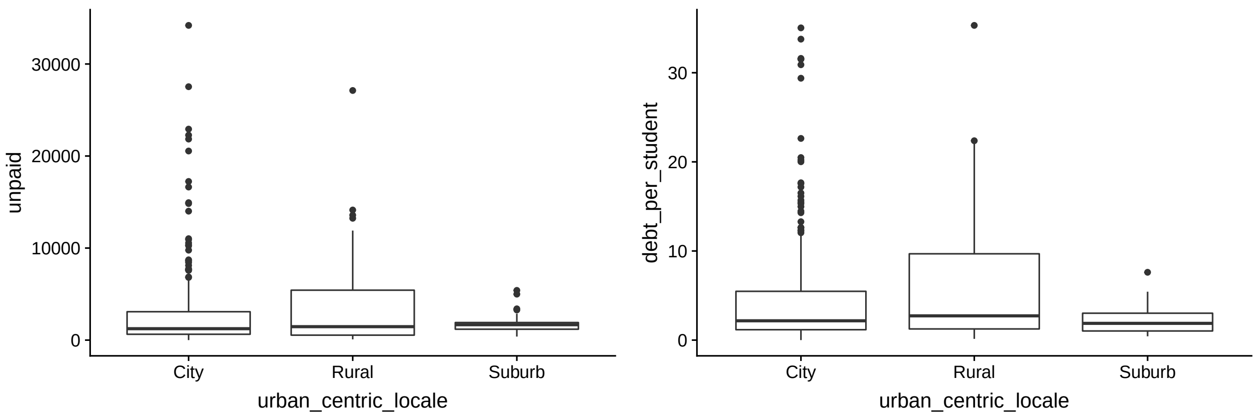
Schools with CEP status generally had a large amount of debt until the 2014-15 school year, when they gained CEP status. That being said, other schools with similar amounts of debt did not gain CEP status.

Demographics and need

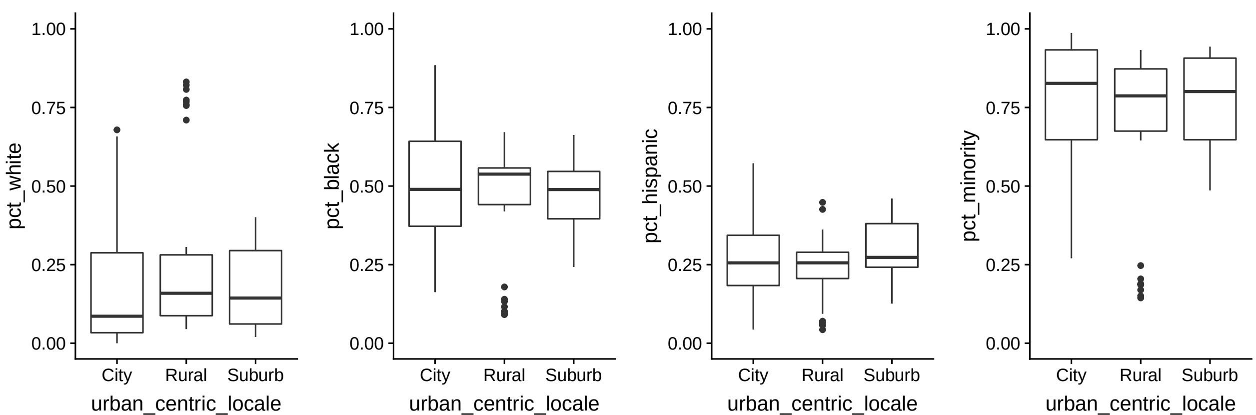


We can also see that race is a proxy for need. The percentage of black and hispanic schools have a strong, positive correlation with the percentage of students on free and reduced price lunch. For white students, this relationship is strong and negative.

Locale and debt, race



There don't seem to be major differences in debt by location.

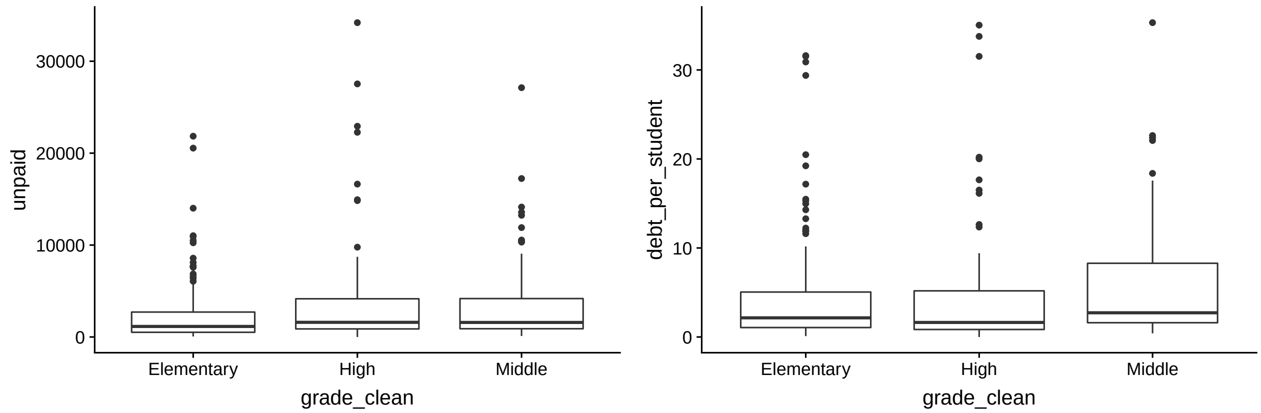


On average, we also don't see any major differences in race by locale. However, there is a subset of rural schools that seem to primarily be white.

Debt and grade

```
## # A tibble: 3 x 3
##   grade_clean mean_debt mean_dps
##   <chr>           <dbl>    <dbl>
## 1 Elementary      2181.     4.10
## 2 High            4016.     4.94
## 3 Middle          3541.     5.70
```

On average, high schools have the most total debt. However, middle schools, on average, have the most debt per student.



CEP Status and Grade Level

```
## # A tibble: 3 x 2
##   grade_clean cep_schools
##   <chr>             <int>
## 1 Elementary           12
## 2 High                  1
## 3 Middle                0
```

Most CEP schools are elementary schools. However, we're also seeing that the schools with the most debt are, on average, middle and high schools.

Regression

Predicting debt per student

I built four linear models to predict debt per student based on various combinations of the percentage of students on free/reduced price lunch, school level, locale and demographics. For these models, I excluded years where CEP schools had CEP status, since their debt would be zero.

```
##
## Call:
## lm(formula = debt_per_student ~ pct_minority + grade_clean +
##     urban_centric_locale, data = df[df$cep == 0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.140 -3.389 -1.280  1.401 28.860
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -2.3122     1.3273  -1.742   0.0824 .
## pct_minority                 8.9830     1.7432   5.153 4.27e-07 ***
## grade_cleanHigh              0.3378     0.8176   0.413   0.6797
## grade_cleanMiddle            0.2843     0.8017   0.355   0.7231
## urban_centric_localeRural    2.2737     0.9865   2.305   0.0218 *
## urban_centric_localeSuburb  -2.0770     1.1831  -1.756   0.0800 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 5.914 on 354 degrees of freedom
## Multiple R-squared:  0.09575,    Adjusted R-squared:  0.08298
## F-statistic: 7.497 on 5 and 354 DF,  p-value: 1.052e-06
```

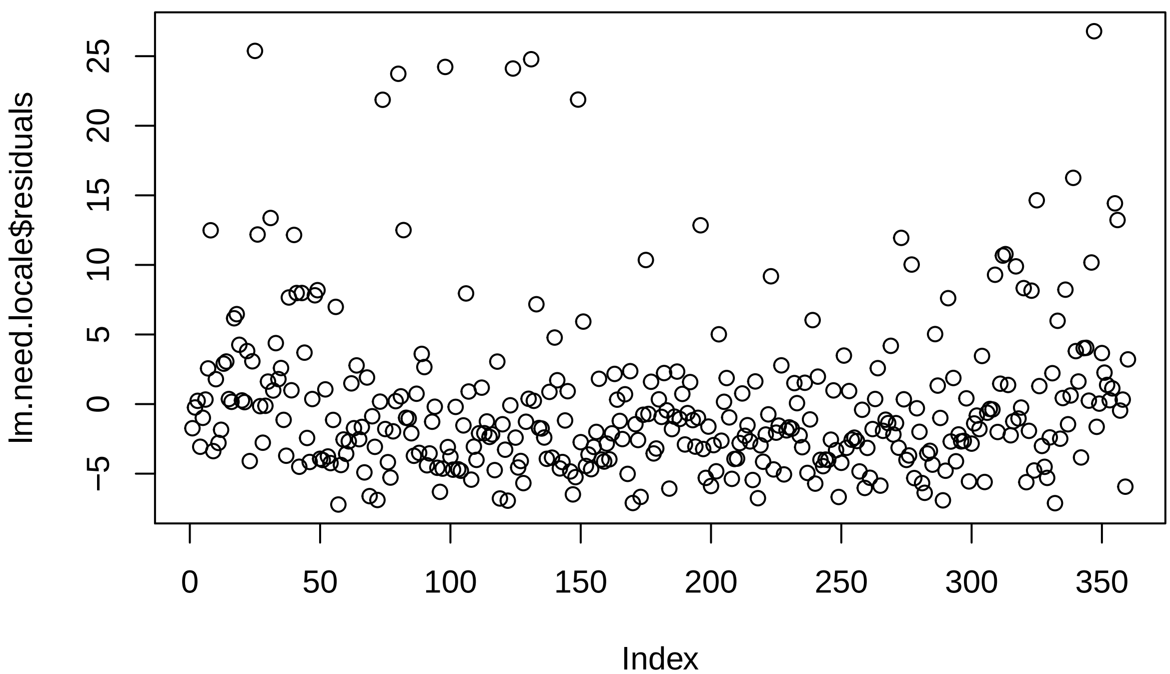
```
##
## Call:
## lm(formula = debt_per_student ~ pct_minority + urban_centric_locale,
##     data = df[df$cep == 0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.303 -3.434 -1.319  1.307 29.025
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -2.3159     1.3227  -1.751   0.0808 .
## pct_minority                 9.1777     1.6882   5.436 1.01e-07 ***
## urban_centric_localeRural    2.2813     0.9516   2.397   0.0170 *
## urban_centric_localeSuburb  -2.1210     1.1683  -1.815   0.0703 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.899 on 356 degrees of freedom
## Multiple R-squared:  0.09515,    Adjusted R-squared:  0.08753
## F-statistic: 12.48 on 3 and 356 DF,  p-value: 8.896e-08
```

```
##
## Call:
## lm(formula = debt_per_student ~ pct_free_reduced + grade_clean +
##     urban_centric_locale, data = df[df$cep == 0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.216 -3.589 -1.418  1.512 26.790
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -3.3171     1.1615  -2.856  0.00455 **
## pct_free_reduced            10.8225     1.5643   6.919 2.15e-11 ***
## grade_cleanHigh              2.1531     0.8007   2.689  0.00750 **
## grade_cleanMiddle            1.3322     0.7628   1.746  0.08162 .
## urban_centric_localeRural    2.2654     0.9569   2.367  0.01845 *
## urban_centric_localeSuburb  -2.2109     1.1515  -1.920  0.05566 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.755 on 354 degrees of freedom
## Multiple R-squared:  0.1437, Adjusted R-squared:  0.1316
## F-statistic: 11.88 on 5 and 354 DF,  p-value: 1.217e-10
```

```
##
## Call:
## lm(formula = debt_per_student ~ pct_free_reduced + grade_clean,
##     data = df[df$cep == 0, ])
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.023 -3.657 -1.698  1.163 28.521
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -2.9827     1.1501  -2.593   0.0099 **
## pct_free_reduced   10.3904     1.5744   6.600  1.5e-10 ***
## grade_cleanHigh     1.8390     0.7997   2.300   0.0220 *
## grade_cleanMiddle   1.8609     0.7526   2.473   0.0139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.82 on 356 degrees of freedom
## Multiple R-squared:  0.1195, Adjusted R-squared:  0.112
## F-statistic:  16.1 on 3 and 356 DF,  p-value: 7.782e-10
```

```
## Analysis of Variance Table
##
## Model 1: debt_per_student ~ pct_minority + grade_clean + urban_centric_locale
## Model 2: debt_per_student ~ pct_minority + urban_centric_locale
## Model 3: debt_per_student ~ pct_free_reduced + grade_clean + urban_centric_locale
## Model 4: debt_per_student ~ pct_free_reduced + grade_clean
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    354 12382
## 2    356 12390 -2     -8.13 0.1162  0.890352
## 3    354 11725  2    664.79 9.5032 9.545e-05 ***
## 4    356 12057 -2   -331.94 4.7451  0.009254 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



None of my models for predicting debt per student explained a large amount of the variance in the data, but I was able to single out the best of my models. The percentage of students who receive free or reduced price lunch is a major predictor of debt per student. For every one percent increase in students who receive free/reduced price lunch, the average debt per student increases by $0.11. Based on our EDA, we also know that schools with a high percentage of free/reduced lunch students tend to be mostly black and hispanic. Our model also found that high schools have $2.15 more debt per student than elementary schools. For middle schools, that value is $1.33. This is interesting, since most CEP schools are actually elementary schools. It also appears that rural schools tend to have more debt per student than urban schools, and suburban schools have less debt per student.

Predicting CEP schools

In this section of my analysis, I try to predict whether a school has CEP status or not. I do this using a logistic regression. The idea here is simple: If I can build a logistic regression model and choose a threshold that accurately predicts the CEP schools, the false positives are, in essence, schools that should also have CEP status. This is an important question: Why do some schools have CEP status while others don't? Especially if there are schools with similar levels of need and debt per student? For my model, I am excluding years where schools had CEP status, and my outcome is `cep_long`, or the longitudinal indicator of whether a school has gained CEP status at any point in time.

```
##
## Call:
## glm(formula = cep_long ~ debt_per_student + pct_free_reduced +
##     grade_clean, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.05996  -0.24601  -0.02768   0.15153   2.74571
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -11.39940    1.68463  -6.767 1.32e-11 ***
## debt_per_student     0.07651    0.05471   1.398   0.1620
## pct_free_reduced    13.76221    2.07021   6.648 2.98e-11 ***
## grade_cleanHigh     -2.03832    1.03681  -1.966   0.0493 *
## grade_cleanMiddle  -19.41383 1190.56950  -0.016   0.9870
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 287.727  on 251  degrees of freedom
## Residual deviance:  98.011  on 247  degrees of freedom
## AIC: 108.01
##
## Number of Fisher Scoring iterations: 18
```

```
##
## Call:
## glm(formula = cep_long ~ debt_per_student + pct_minority + grade_clean,
##     family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.84243  -0.01228  -0.00009   0.01085   2.39798
##
```

```
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -39.32792    8.91767  -4.410 1.03e-05 ***
## debt_per_student     0.22056    0.06448   3.420 0.000625 ***
## pct_minority        44.51723    9.97928   4.461 8.16e-06 ***
## grade_cleanHigh     -5.67267    1.17918  -4.811 1.50e-06 ***
## grade_cleanMiddle  -23.75768 1864.75620  -0.013 0.989835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 287.727  on 251  degrees of freedom
## Residual deviance:  58.167  on 247  degrees of freedom
## AIC: 68.167
##
## Number of Fisher Scoring iterations: 19
```

```
## Analysis of Deviance Table
##
## Model 1: cep_long ~ debt_per_student + pct_free_reduced + grade_clean
## Model 2: cep_long ~ debt_per_student + pct_minority + grade_clean
##   Resid. Df Resid. Dev Df Deviance
## 1       247     98.011
## 2       247     58.167  0   39.844
```

I'll stick with my second model, since they are both roughly good predictors of whether a school gains CEP status.

```
##       (Intercept)  debt_per_student      pct_minority   grade_cleanHigh
##          8.32e-18          1.25e+00          2.16e+19          3.44e-03
## grade_cleanMiddle
##          4.81e-11
```

```
## pct_minority
##         1.56
```

For every additional dollar of debt per student, schools are 1.24 times more likely to receive CEP status. For every additional percentage point of minority students, schools are 1.56 times more likely to gain CEP status. We also see that high schools are 0.003 times less likely to be assigned CEP status. The intercept for middle schools is not significant.

Predictions

```
##    school_no           school_name debt_per_student pct_free_reduced
## 1        304              Bethesda       0.05263158        0.5593985
## 2        308                Burton       3.32753623        0.9826087
## 3        310               Eastway       0.01875862        0.8913793
## 4        315            Eno Valley       0.38745552        0.9590747
## 5        320                 Glenn       0.02844950        0.5945946
## 6        328                  Holt       0.68203675        0.9264931
## 7        339   Lakewood Elementary       0.03773784        0.7209302
## 8        341              Lakeview       3.57128713        1.0000000
## 9        344   Fayetteville Street       0.12050360        0.7410072
## 10       352          Merrick Moore       0.02340741        0.6296296
## 11       360              Oak Grove       2.96669394        0.9116203
```

```
## 12       367             R N Harris       0.02248538        0.6695906
## 13       368               Southern      16.50677539        0.7064492
## 14       374           C C Spaulding       0.00000000        0.9924528
## 15       388 W G Pearson Elementary       0.75784351        0.9656489
## 16       400               Y E Smith       0.00000000        0.8293963
##    grade_clean cep predict_ind diff
## 1   Elementary   1           1   -1
## 2   Elementary   0           1    0
## 3   Elementary   1           1   -1
## 4   Elementary   0           1    0
## 5   Elementary   1           1   -1
## 6   Elementary   0           1    0
## 7   Elementary   1           1   -1
## 8         High   1           0   -2
## 9   Elementary   1           1   -1
## 10  Elementary   1           1   -1
## 11  Elementary   0           1    0
## 12  Elementary   1           1   -1
## 13        High   0           1    0
## 14  Elementary   1           1   -1
## 15  Elementary   0           1    0
## 16  Elementary   1           1   -1
```

Based on my model (which isn't the best out there), there are 6 schools that were not assigned CEP status even though they have similar assets.

K-means models

In this section, I'll try to group the schools by high and low need through a k-means clustering algorithm.

```
##  [1] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1
## [36] 1 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1
```

```
## [1] 209819186
```

NOTES:

- CEP provisions began in 2014-15 in NC
- data collected at different times so numbers won't match perfectly
- 353 is housed in Durham Tech, which complicates some of the data
- For the most part, aside from missing data, it seems like schools with no debt are CEP schools. We can't get CEP status starting in 2010-11, since those were pilot years.
- CEP schools for 2017-18 all have > 90% free/reduced, setting to 100% (doesn't affect analysis since excluded from regression)

REPORTING QUESTIONS: - how do students get assigned to schools? any diversity initiatives? - does the district track when debt is payed off? - can a student graduate with debt? are they barred from anything? what are consequences besides food?