

# GRF for Saga/Match Treatment Heterogeneity

This document presents the current GRF (Generalized Random Forests) analysis for treatment heterogeneity. We present results for the following outcomes:

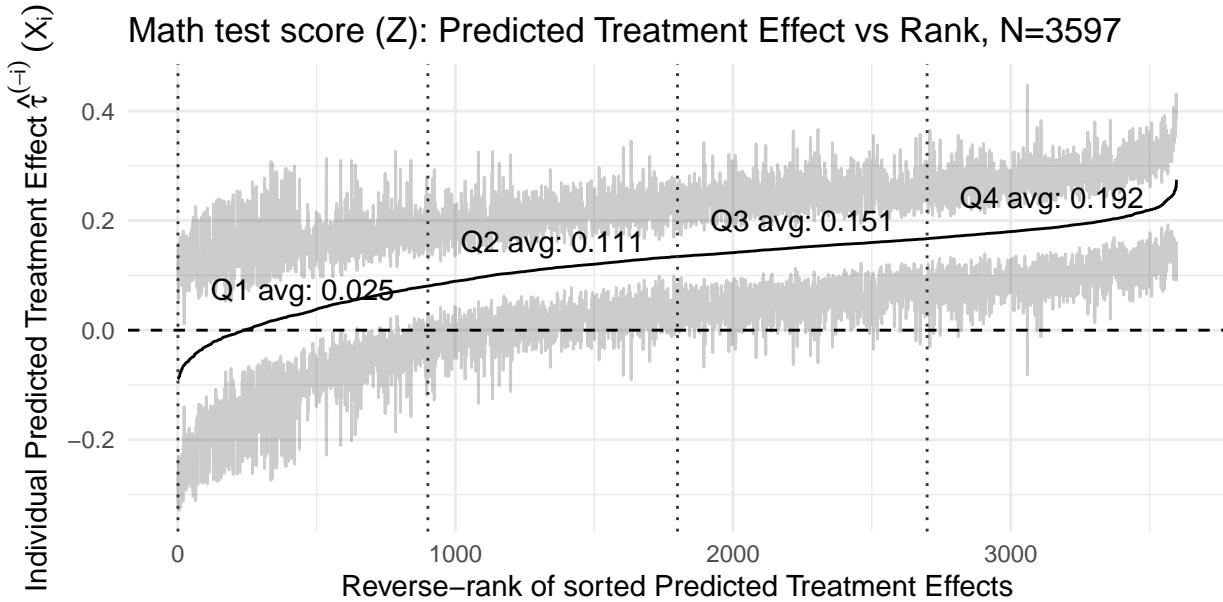
- Math Test Scores
- Math Class Failures
- Math GPA
- Reading Test Scores
- Non-math GPA
- Graduated on-time
- Ever graduated
- Participation in Saga tutoring in year 2
- 11th Grade Math GPA

We use the most recent release of the GRF package by Tibshirani, Athey, et. al. We try to follow the example of Jon Davis and Sara Heller wherever possible, but in the interceding years there has been some updates in the underlying package, so there are some differences. Some implementation details:

- We use a training sample split (`sample.fraction`) of 0.50 when building each tree in each forest
  - The package notes “when variance estimates are requested, `sample.fraction` cannot be greater than 0.5”
- For each outcome of interest, we grow 100,000 trees to make each causal forest
- Following Davis + Heller, we “adjust for differences in treatment probabilities [by] using inverse probability weights throughout the procedure” (following their calculation)
- Jon Davis + Sara Heller dealt with missingness in covariates by imputing block means and including missingness dummies
  - Since then, the underlying code has been updated with its own methods to deal with missingness (treating NAs as their own category during each split)
- The package now supports ‘clustering’, so we cluster observations at the individual level to account for multiple observations of students randomized multiple times in study 2 (ensuring the same student can’t be in both test/train splits when fitting each individual tree)
  - We now exclude all students with duplicate observations.

We use all covariates from our main analyses (except for randomization block). These include gender, age, learning disability, free lunch recipient, race, baseline grade level, GPA, baseline test performance (and within-baseline-school math test decile), days absent from school, disciplinary incidents, including suspensions, and arrests.

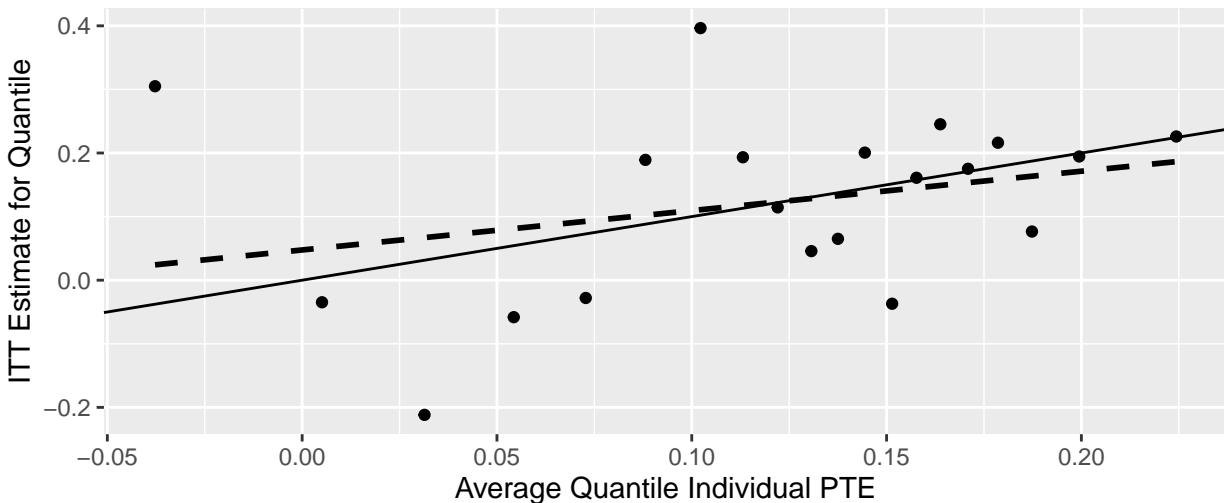
## Outcome: Math test score (Z)



above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitors”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.119 (0.026).<sup>1</sup>

### Calibration Plot: mathxil\_z\_post1\_np

PTE x Treatment interaction estimate is 0.606 (0.396), with a p-value of 0.126



<sup>1</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.002 and a ‘differential forest prediction’ (DFP) coefficient of 0.495, with a corresponding DFP p-value of 0.113. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 1: Average Treatment Effects (overall and for subsamples) for Math test score (Z)

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.119 +/- 0.05	0.026
Individual PTE Quartile 4	0.151 +/- 0.092	0.047
Individual PTE Quartile 3	0.116 +/- 0.096	0.049
Individual PTE Quartile 2	0.242 +/- 0.104	0.053
Individual PTE Quartile 1	-0.03 +/- 0.11	0.056
Top 2 PTE Quartiles	0.133 +/- 0.066	0.034
Bottom 2 PTE Quartiles	0.105 +/- 0.076	0.039
Bottom 3 PTE Quartiles	0.109 +/- 0.06	0.030

Table 2: Differences between subsample average treatment effects: Math test score (Z)

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.068, 0.151]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.073, 0.129]
Quartile 4	Quartile 1	[0.038, 0.324]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.591979882090698

Tuned parameters:

```
sample.fraction: 0.366547763661947
mtry: 5
honesty.fraction: 0.659975905413739
honesty.prune.leaves: 1
alpha: 0.00110271532321349
imbalance.penalty: 0.36186038762885
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0502,0.14] 0.5926037
(0.14,0.232] 0.5926995
(0.232,0.314] 0.5929056
(0.314,0.411] 0.5929465
(0.411,0.5] 0.5933660
```

```
mtry      error
[1,6] 0.5933688
(6,11] 0.5926046
(11,17] 0.5927569
(17,22] 0.5928092
(22,26] 0.5929192
```

```
honesty.fraction      error
[0.5,0.558] 0.5928889
(0.558,0.622] 0.5928833
(0.622,0.678] 0.5928196
(0.678,0.738] 0.5929654
```

```

(0.738,0.8] 0.5929642

honesty.prune.leaves      error
  0 0.5929064
  1 0.5929021

      alpha      error
[7.01e-06,0.049] 0.5925473
(0.049,0.0994] 0.5926435
(0.0994,0.147] 0.5928430
(0.147,0.198] 0.5930911
(0.198,0.25] 0.5933965

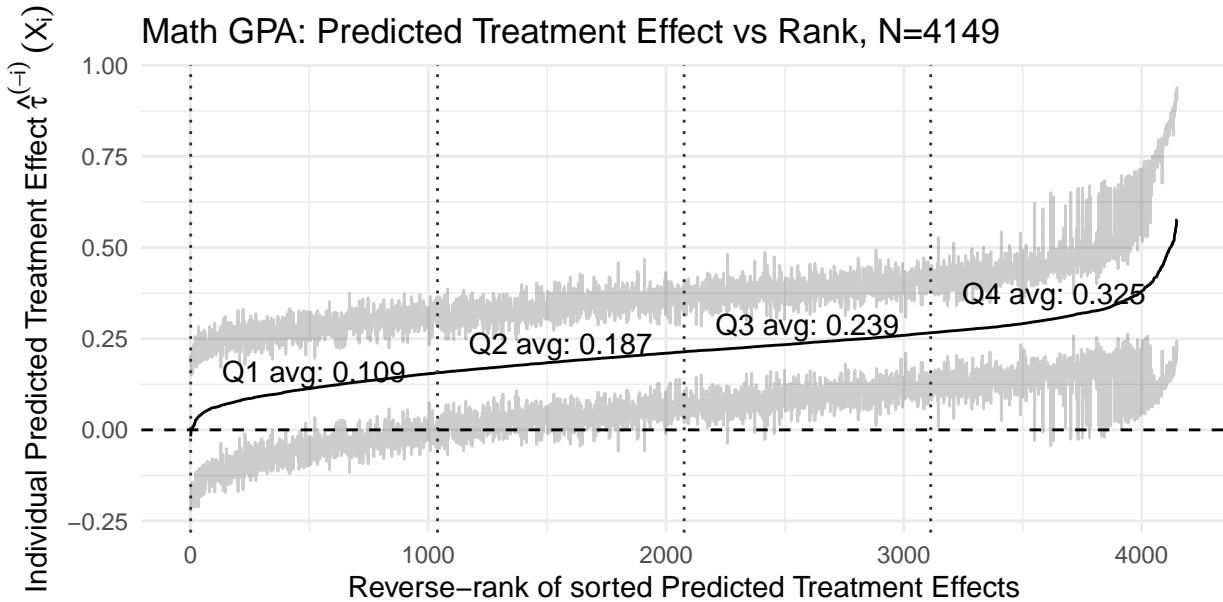
imbalance.penalty      error
[0.000242,0.224] 0.5928791
(0.224,0.511] 0.5929038
(0.511,0.913] 0.5928417
(0.913,1.58] 0.5929376
(1.58,8.06] 0.5929592

```

Table 3: Summary table by Quartile of Predicted Treatment Effects on Math test score (Z)

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.025	0.111	0.151	0.192
N	900.000	899.000	899.000	899.000
Age	14.836	14.594	14.446	14.313
Female	0.151	0.141	0.147	0.197
Has IEP	0.348	0.236	0.080	0.016
Has Free/Reduced Lunch	0.892	0.903	0.889	0.888
Black	0.589	0.560	0.513	0.378
Hispanic	0.377	0.417	0.449	0.542
Other Race	0.034	0.023	0.038	0.080
In 9th Grade	0.686	0.735	0.719	0.775
In 10th Grade	0.308	0.265	0.281	0.225
Baseline GPA	1.729	2.294	2.659	2.771
Num. A's	2.813	5.849	8.756	11.972
Num. B's	3.696	7.826	9.345	8.958
Num. C's	7.109	9.954	8.375	7.507
Num. D's	4.469	3.595	2.517	2.506
Num. F's	4.293	2.432	1.434	1.330
Missing Baseline GPA/Grades	0.171	0.037	0.026	0.042
Days Absent	22.417	14.558	10.633	8.372
Missing Attendance Data	0.133	0.000	0.000	0.000
Math Test Score (Z)	-1.243	-0.079	0.416	0.587
Reading Test Score (Z)	-0.891	0.024	0.364	0.327
Missing Math Test	0.296	0.022	0.007	0.004
Missing Reading Test	0.297	0.026	0.010	0.008
Out-of-School Suspensions	2.190	1.060	0.366	0.116
Disciplinary Incidents	1.721	0.836	0.318	0.125
Any Arrests at Baseline	0.174	0.115	0.081	0.041
Arrests: Violent Crime	0.114	0.069	0.038	0.011
Arrests: Property Crime	0.043	0.033	0.027	0.014
Arrests: Drug Crime	0.040	0.020	0.012	0.004
Math Score - Decile in Previous School	2.412	5.468	7.086	7.437
In Study 2	0.438	0.477	0.486	0.556
Participated in Year 1 of Study	0.293	0.269	0.249	0.258

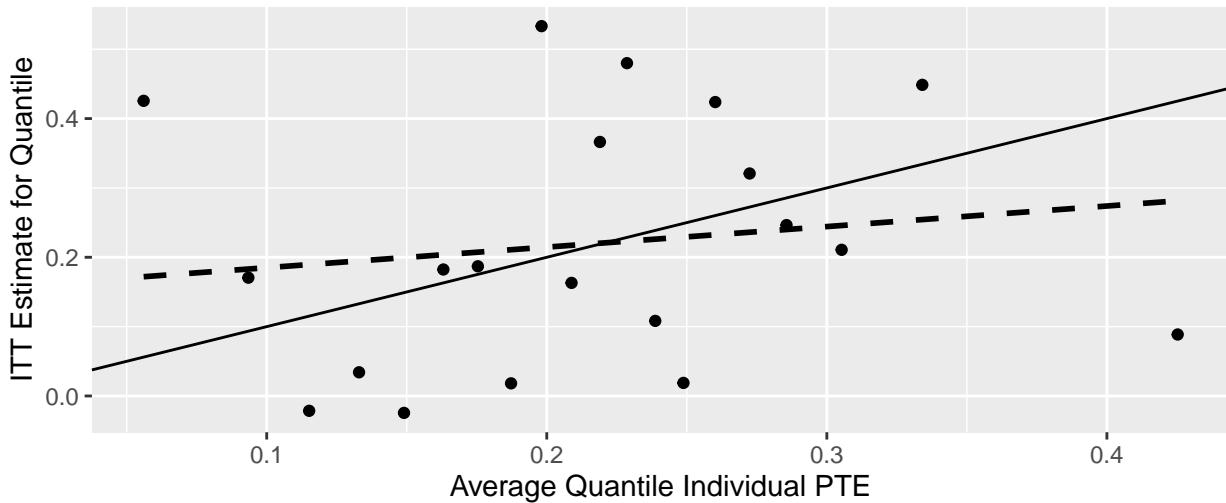
## Outcome: Math GPA



The above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.216 (0.029).<sup>2</sup>

### Calibration Plot: mathgpa\_post1

PTE x Treatment interaction estimate is 0.211 (0.344), with a p-value of 0.54



<sup>2</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.005 and a ‘differential forest prediction’ (DFP) coefficient of 0.269, with a corresponding DFP p-value of 0.236. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 4: Average Treatment Effects (overall and for subsamples) for Math GPA

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.216 +/- 0.057	0.029
Individual PTE Quartile 4	0.251 +/- 0.119	0.061
Individual PTE Quartile 3	0.295 +/- 0.111	0.056
Individual PTE Quartile 2	0.22 +/- 0.115	0.058
Individual PTE Quartile 1	0.096 +/- 0.114	0.058
Top 2 PTE Quartiles	0.273 +/- 0.081	0.041
Bottom 2 PTE Quartiles	0.159 +/- 0.081	0.041
Bottom 3 PTE Quartiles	0.204 +/- 0.065	0.033

Table 5: Differences between subsample average treatment effects: Math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.089, 0.183]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[0, 0.229]
Quartile 4	Quartile 1	[-0.01, 0.32]

Tuning status: default.

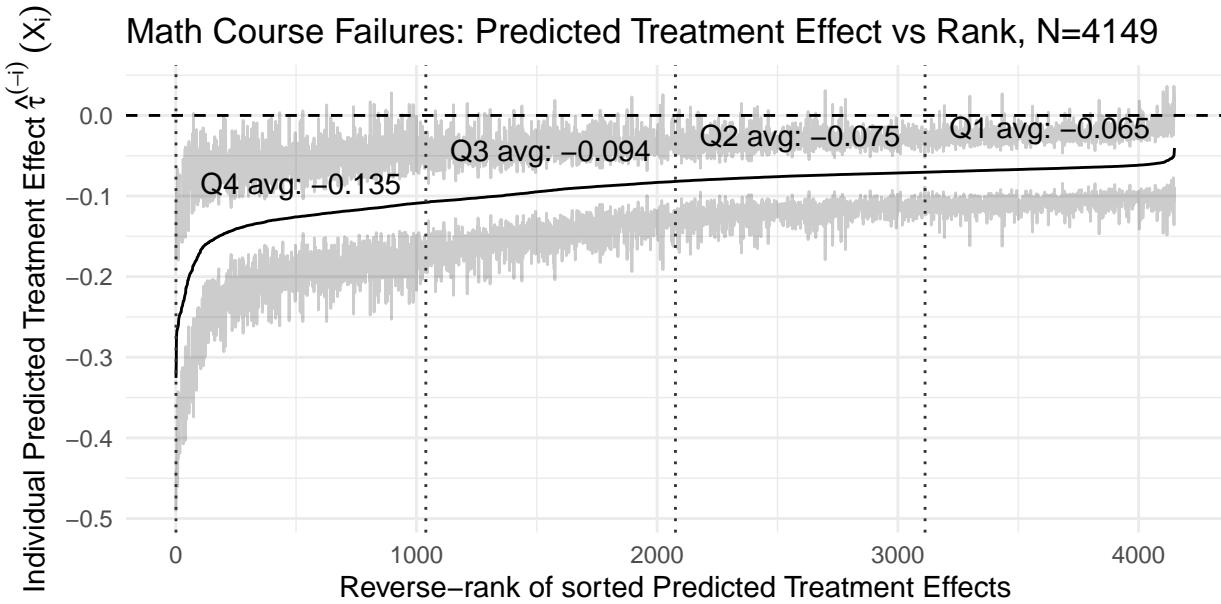
This indicates tuning was attempted. However, we could not find parameters that were expected to perform best.

```
sample.fraction: 0.5
mtry: 26
honesty.fraction: 0.5
honesty.prune.leaves: TRUE
alpha: 0.05
imbalance.penalty: 0
```

Table 6: Summary table by Quartile of Predicted Treatment Effects on Math GPA

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.109	0.187	0.239	0.325
N	1038.000	1037.000	1037.000	1037.000
Age	14.287	14.545	14.882	14.876
Female	0.178	0.195	0.131	0.120
Has IEP	0.197	0.172	0.188	0.134
Has Free/Reduced Lunch	0.940	0.912	0.891	0.810
Black	0.537	0.563	0.528	0.447
Hispanic	0.416	0.408	0.436	0.486
Other Race	0.047	0.029	0.036	0.067
In 9th Grade	0.984	0.788	0.554	0.474
In 10th Grade	0.016	0.206	0.435	0.521
Baseline GPA	2.345	2.249	2.199	2.220
Num. A's	9.791	7.348	5.853	3.057
Num. B's	8.427	7.443	6.440	4.566
Num. C's	11.441	8.539	6.575	4.331
Num. D's	5.707	3.253	2.535	1.846
Num. F's	3.554	3.837	2.691	0.983
Missing Baseline GPA/Grades	0.000	0.000	0.012	0.308
Days Absent	14.623	19.885	19.867	14.924
Missing Attendance Data	0.000	0.000	0.000	0.178
Math Test Score (Z)	-0.127	0.050	-0.087	0.029
Reading Test Score (Z)	-0.036	0.064	-0.086	-0.072
Missing Math Test	0.067	0.121	0.107	0.220
Missing Reading Test	0.072	0.126	0.109	0.221
Out-of-School Suspensions	1.107	1.902	1.568	0.683
Disciplinary Incidents	0.706	1.390	1.265	0.637
Any Arrests at Baseline	0.116	0.200	0.184	0.120
Arrests: Violent Crime	0.085	0.169	0.116	0.067
Arrests: Property Crime	0.051	0.086	0.077	0.044
Arrests: Drug Crime	0.028	0.057	0.093	0.029
Math Score - Decile in Previous School	5.386	5.924	5.691	6.153
In Study 2	0.576	0.547	0.396	0.361
Participated in Year 1 of Study	0.277	0.260	0.254	0.257

## Outcome: Math Course Failures

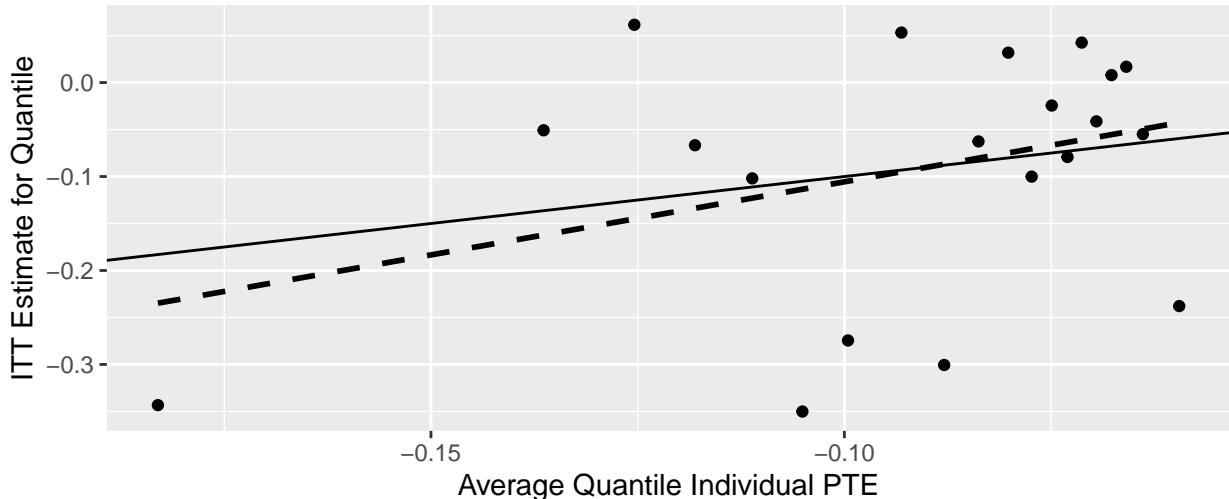


The

above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ 's  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of -0.092

### Calibration Plot: mathfail\_post1

PTE x Treatment interaction estimate is 1.881 (0.63), with a p-value of 0.003



$(0.02)^3$

<sup>3</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 0.998 and a ‘differential forest prediction’ (DFP) coefficient of 2.085, with a corresponding DFP p-value of 0.025. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 7: Average Treatment Effects (overall and for subsamples) for Math Course Failures

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	-0.092 +/- 0.039	0.020
Individual PTE Quartile 4	-0.111 +/- 0.104	0.053
Individual PTE Quartile 3	-0.165 +/- 0.083	0.042
Individual PTE Quartile 2	-0.02 +/- 0.06	0.030
Individual PTE Quartile 1	-0.075 +/- 0.054	0.028
Top 2 PTE Quartiles	-0.138 +/- 0.067	0.034
Bottom 2 PTE Quartiles	-0.047 +/- 0.04	0.021
Bottom 3 PTE Quartiles	-0.086 +/- 0.039	0.020

Table 8: Differences between subsample average treatment effects: Math Course Failures

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.136, 0.086]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.169, -0.013]
Quartile 4	Quartile 1	[-0.153, 0.082]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.407364548566824

Tuned parameters:

```
sample.fraction: 0.489287406601943
mtry: 1
honesty.fraction: 0.755155843077227
honesty.prune.leaves: 0
alpha: 0.0113845444284379
imbalance.penalty: 1.07544889189395
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0504,0.139] 0.4084186
(0.139,0.225] 0.4087084
(0.225,0.318] 0.4089871
(0.318,0.403] 0.4091690
(0.403,0.5] 0.4091509
```

```
mtry      error
[1,5] 0.4084001
(5,11] 0.4086976
(11,16] 0.4090433
(16,21] 0.4091692
(21,26] 0.4091952
```

```
honesty.fraction      error
[0.5,0.561] 0.4087638
(0.561,0.619] 0.4088431
(0.619,0.681] 0.4089148
(0.681,0.739] 0.4089867
```

```

(0.739,0.8] 0.4089257

honesty.prune.leaves      error
  0 0.4088662
  1 0.4089069

      alpha      error
[0.000148,0.0509] 0.4086852
(0.0509,0.101] 0.4087932
(0.101,0.15] 0.4089253
(0.15,0.202] 0.4089970
(0.202,0.25] 0.4090334

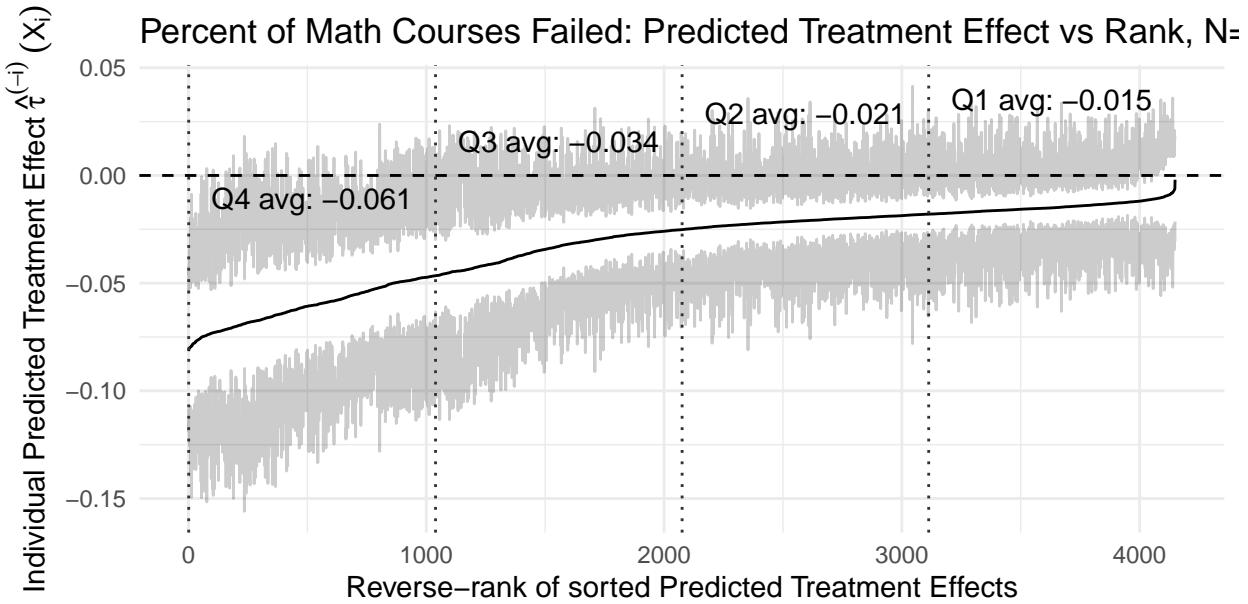
imbalance.penalty      error
[0.000988,0.234] 0.4089573
(0.234,0.523] 0.4089082
(0.523,0.899] 0.4089273
(0.899,1.57] 0.4088288
(1.57,8.8] 0.4088125

```

Table 9: Summary table by Quartile of Predicted Treatment Effects on Math Course Failures

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.065	-0.075	-0.094	-0.135
N	1037.000	1037.000	1037.000	1038.000
Age	14.356	14.368	14.804	15.061
Female	0.216	0.187	0.132	0.089
Has IEP	0.203	0.167	0.177	0.143
Has Free/Reduced Lunch	0.987	0.896	0.805	0.865
Black	0.602	0.449	0.559	0.465
Hispanic	0.398	0.526	0.334	0.488
Other Race	0.000	0.025	0.107	0.046
In 9th Grade	0.927	0.769	0.522	0.584
In 10th Grade	0.073	0.230	0.473	0.401
Baseline GPA	2.833	2.710	2.178	1.089
Num. A's	12.018	9.142	3.869	1.029
Num. B's	11.797	9.044	4.443	1.597
Num. C's	10.712	9.590	6.767	3.825
Num. D's	2.503	2.828	3.399	4.611
Num. F's	1.089	1.446	2.041	6.487
Missing Baseline GPA/Grades	0.000	0.000	0.144	0.175
Days Absent	9.768	9.907	17.723	33.947
Missing Attendance Data	0.000	0.000	0.082	0.096
Math Test Score (Z)	0.393	-0.032	-0.201	-0.443
Reading Test Score (Z)	0.310	0.003	-0.158	-0.401
Missing Math Test	0.015	0.043	0.183	0.273
Missing Reading Test	0.018	0.043	0.189	0.277
Out-of-School Suspensions	0.354	0.366	1.145	3.723
Disciplinary Incidents	0.234	0.281	0.924	2.788
Any Arrests at Baseline	0.109	0.060	0.169	0.281
Arrests: Violent Crime	0.064	0.032	0.126	0.214
Arrests: Property Crime	0.030	0.020	0.068	0.141
Arrests: Drug Crime	0.021	0.009	0.046	0.130
Math Score - Decile in Previous School	6.907	5.625	5.379	4.792
In Study 2	0.609	0.529	0.417	0.325
Participated in Year 1 of Study	0.254	0.264	0.282	0.249

## Outcome: Percent of Math Courses Failed

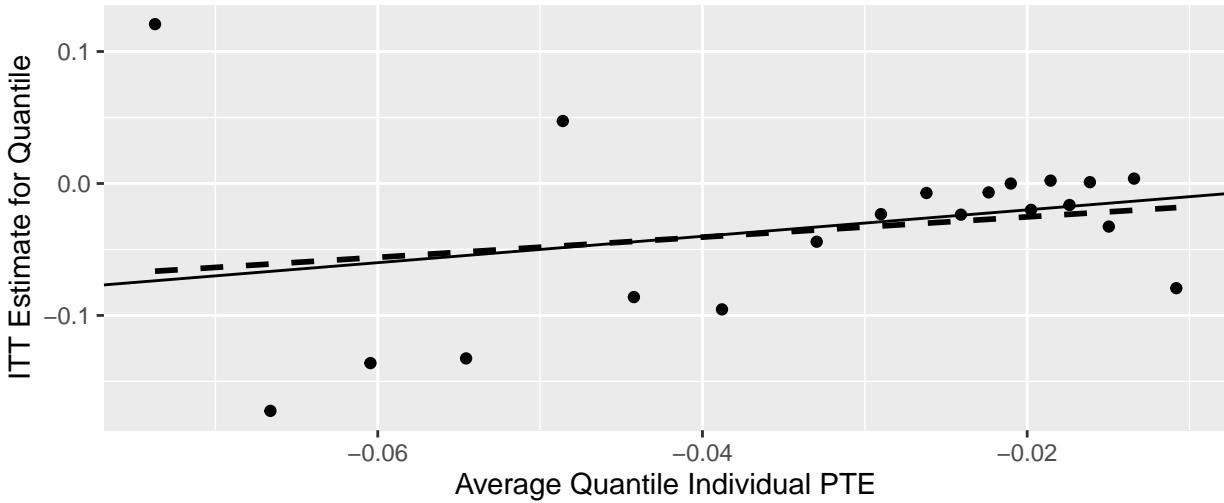


The

above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of -0.033 (0.009).<sup>4</sup>

### Calibration Plot: mathfailpercent\_post1

PTE x Treatment interaction estimate is 0.656 (0.477), with a p-value of 0.17



<sup>4</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.008 and a ‘differential forest prediction’ (DFP) coefficient of 0.697, with a corresponding DFP p-value of 0.122. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 10: Average Treatment Effects (overall and for subsamples) for Percent of Math Courses Failed

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	-0.033 +/- 0.017	0.009
Individual PTE Quartile 4	-0.056 +/- 0.047	0.024
Individual PTE Quartile 3	-0.045 +/- 0.037	0.019
Individual PTE Quartile 2	-0.006 +/- 0.025	0.013
Individual PTE Quartile 1	-0.023 +/- 0.025	0.013
Top 2 PTE Quartiles	-0.051 +/- 0.03	0.015
Bottom 2 PTE Quartiles	-0.015 +/- 0.018	0.009
Bottom 3 PTE Quartiles	-0.025 +/- 0.017	0.009

Table 11: Differences between subsample average treatment effects: Percent of Math Courses Failed

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.081, 0.019]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.071, -0.001]
Quartile 4	Quartile 1	[-0.086, 0.021]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.0821468429015715

Tuned parameters:

```
sample.fraction: 0.0847166027291678
mtry: 11
honesty.fraction: 0.667728955391794
honesty.prune.leaves: 1
alpha: 0.0131427262094803
imbalance.penalty: 0.131784388001513
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0501,0.14] 0.08211290
(0.14,0.232] 0.08212401
(0.232,0.322] 0.08213277
(0.322,0.413] 0.08215937
(0.413,0.5] 0.08220840
```

```
mtry      error
[1,6] 0.08213065
(6,11] 0.08212361
(11,16] 0.08214744
(16,21] 0.08216162
(21,26] 0.08218039
```

```
honesty.fraction      error
[0.5,0.557] 0.08213379
(0.557,0.618] 0.08214286
(0.618,0.676] 0.08214914
(0.676,0.738] 0.08215017
```

```
(0.738,0.8] 0.08216150

honesty.prune.leaves      error
  0 0.08214654
  1 0.08214844

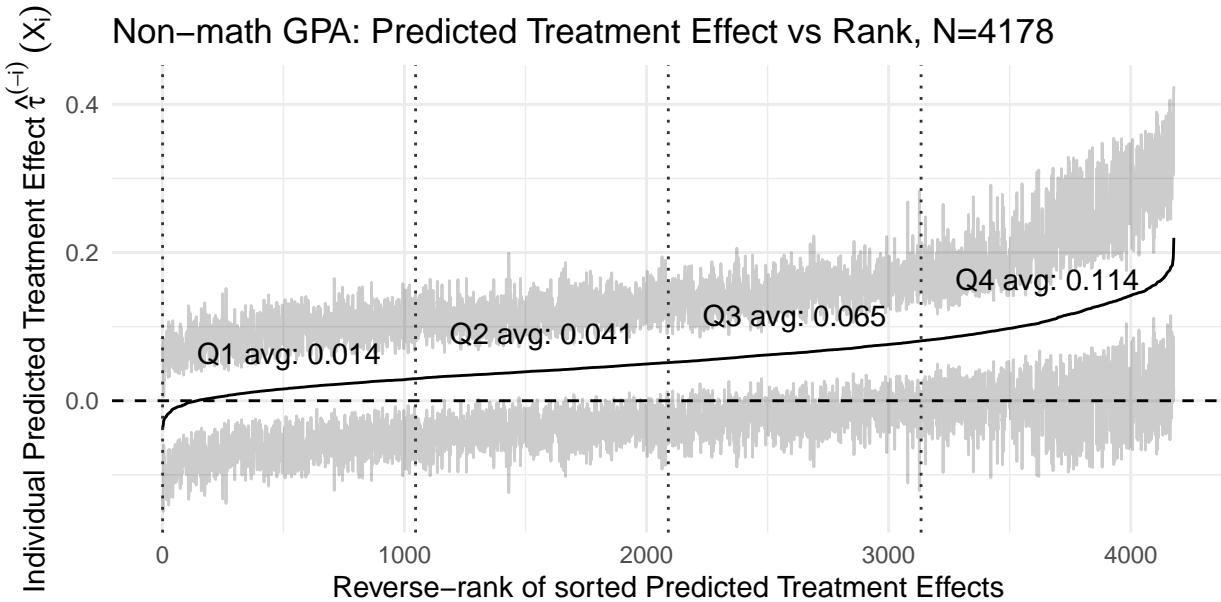
      alpha      error
[2.45e-05,0.0529] 0.08211235
(0.0529,0.103] 0.08212708
(0.103,0.155] 0.08214683
(0.155,0.202] 0.08216525
(0.202,0.25] 0.08218594

imbalance.penalty      error
[0.000186,0.224] 0.08217288
(0.224,0.514] 0.08215344
(0.514,0.958] 0.08214678
(0.958,1.64] 0.08213470
(1.64,8.32] 0.08212966
```

Table 12: Summary table by Quartile of Predicted Treatment Effects on Percent of Math Courses Failed

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.015	-0.021	-0.034	-0.061
N	1037.000	1037.000	1037.000	1038.000
Age	14.200	14.442	14.770	15.177
Female	0.195	0.200	0.135	0.094
Has IEP	0.125	0.204	0.219	0.142
Has Free/Reduced Lunch	0.940	0.895	0.844	0.875
Black	0.501	0.468	0.567	0.539
Hispanic	0.450	0.495	0.381	0.420
Other Race	0.048	0.038	0.052	0.040
In 9th Grade	0.993	0.774	0.566	0.467
In 10th Grade	0.007	0.225	0.429	0.517
Baseline GPA	2.780	2.761	2.195	1.026
Num. A's	12.713	9.406	3.426	0.513
Num. B's	11.637	9.735	4.502	1.008
Num. C's	11.007	9.687	7.190	3.011
Num. D's	3.125	2.413	3.643	4.161
Num. F's	1.391	1.388	2.769	5.516
Missing Baseline GPA/Grades	0.000	0.000	0.110	0.209
Days Absent	9.958	9.489	15.815	35.753
Missing Attendance Data	0.000	0.000	0.099	0.079
Math Test Score (Z)	0.534	-0.141	-0.328	-0.335
Reading Test Score (Z)	0.366	-0.057	-0.240	-0.293
Missing Math Test	0.027	0.040	0.190	0.258
Missing Reading Test	0.031	0.041	0.194	0.262
Out-of-School Suspensions	0.171	0.383	1.543	3.467
Disciplinary Incidents	0.169	0.289	1.018	2.723
Any Arrests at Baseline	0.075	0.071	0.157	0.315
Arrests: Violent Crime	0.029	0.039	0.129	0.239
Arrests: Property Crime	0.028	0.017	0.064	0.149
Arrests: Drug Crime	0.015	0.008	0.033	0.150
Math Score - Decile in Previous School	7.300	5.273	5.023	5.175
In Study 2	0.618	0.544	0.416	0.303
Participated in Year 1 of Study	0.270	0.257	0.281	0.240

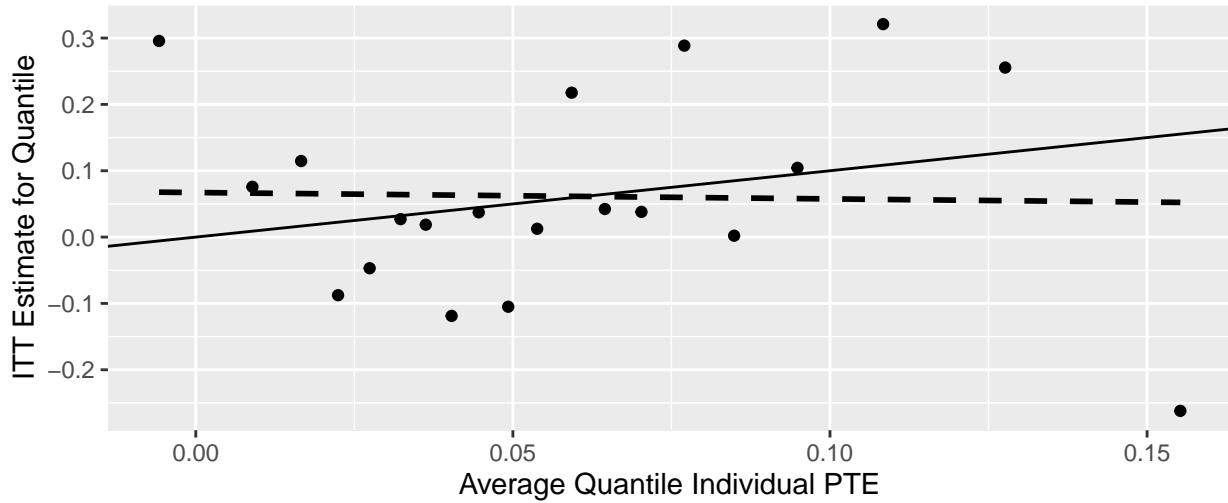
## Outcome: Non-math GPA



above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.058 (0.023).<sup>5</sup>

### Calibration Plot: nonmathgpa\_post1

PTE x Treatment interaction estimate is  $-0.088$  ( $0.578$ ), with a p-value of  $0.879$



<sup>5</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.004 and a ‘differential forest prediction’ (DFP) coefficient of -0.316, with a corresponding DFP p-value of 0.687. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 13: Average Treatment Effects (overall and for subsamples) for Non-math GPA

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.058 +/- 0.045	0.023
Individual PTE Quartile 4	0.067 +/- 0.095	0.049
Individual PTE Quartile 3	0.114 +/- 0.088	0.045
Individual PTE Quartile 2	0.012 +/- 0.089	0.045
Individual PTE Quartile 1	0.038 +/- 0.089	0.045
Top 2 PTE Quartiles	0.091 +/- 0.065	0.033
Bottom 2 PTE Quartiles	0.026 +/- 0.063	0.032
Bottom 3 PTE Quartiles	0.055 +/- 0.051	0.026

Table 14: Differences between subsample average treatment effects: Non-math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.096, 0.12]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.025, 0.155]
Quartile 4	Quartile 1	[-0.102, 0.159]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.551018698694469

Tuned parameters:

```
sample.fraction: 0.395819516247138
mtry: 3
honesty.fraction: 0.682501491578296
honesty.prune.leaves: 1
alpha: 0.000407631567213684
imbalance.penalty: 0.216328855012685
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0501,0.144] 0.5524701
(0.144,0.23] 0.5524173
(0.23,0.324] 0.5525011
(0.324,0.413] 0.5527841
(0.413,0.5] 0.5528823
```

```
mtry      error
[1,6] 0.5523652
(6,11] 0.5524343
(11,15] 0.5525693
(15,21] 0.5528132
(21,26] 0.5529342
```

```
honesty.fraction      error
[0.5,0.561] 0.5525246
(0.561,0.622] 0.5525377
(0.622,0.682] 0.5526099
(0.682,0.739] 0.5526433
```

```

(0.739,0.8] 0.5527395

honesty.prune.leaves      error
  0 0.5526249
  1 0.5525980

      alpha      error
[0.000408,0.0521] 0.5524105
(0.0521,0.098] 0.5524349
(0.098,0.149] 0.5526348
(0.149,0.199] 0.5527050
(0.199,0.25] 0.5528697

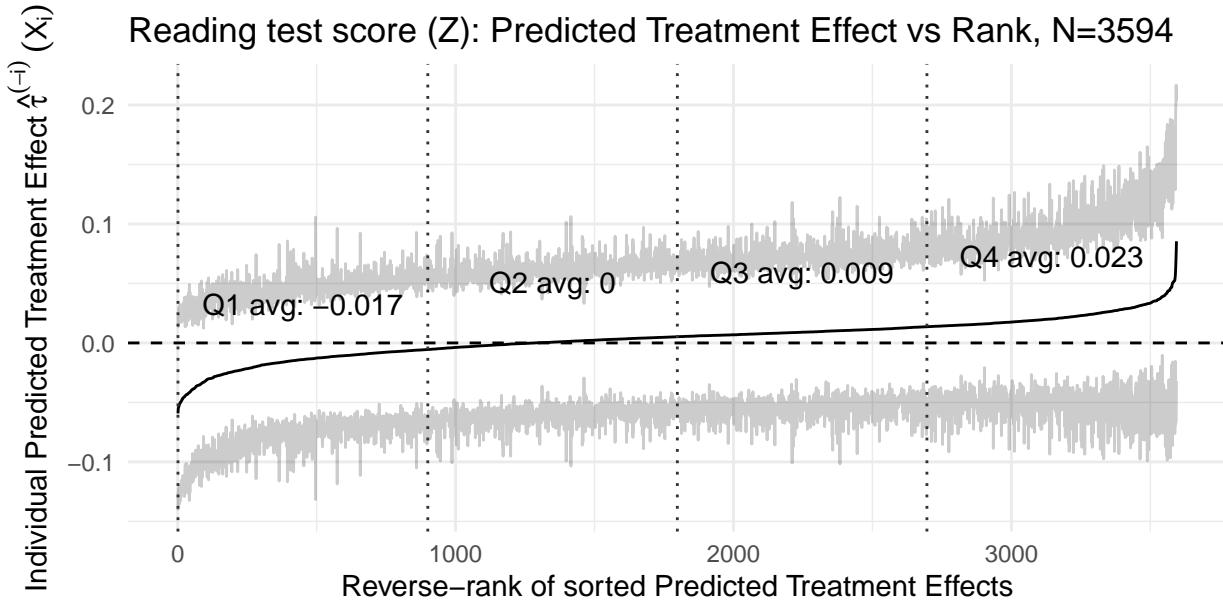
imbalance.penalty      error
[0.000444,0.219] 0.5526718
(0.219,0.514] 0.5526358
(0.514,0.892] 0.5526224
(0.892,1.52] 0.5525901
(1.52,6.83] 0.5525349

```

Table 15: Summary table by Quartile of Predicted Treatment Effects on Non-math GPA

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.014	0.041	0.065	0.114
N	1045.000	1044.000	1044.000	1045.000
Age	14.607	14.550	14.500	14.955
Female	0.086	0.187	0.207	0.144
Has IEP	0.182	0.193	0.199	0.115
Has Free/Reduced Lunch	0.941	0.908	0.868	0.839
Black	0.362	0.513	0.571	0.633
Hispanic	0.618	0.435	0.371	0.320
Other Race	0.020	0.052	0.058	0.047
In 9th Grade	0.681	0.749	0.776	0.590
In 10th Grade	0.319	0.248	0.221	0.393
Baseline GPA	2.405	2.376	2.456	1.607
Num. A's	6.038	7.550	9.093	3.252
Num. B's	8.694	7.966	7.527	2.561
Num. C's	10.706	9.792	7.332	2.956
Num. D's	4.050	3.637	2.883	2.794
Num. F's	1.399	2.416	2.774	4.620
Missing Baseline GPA/Grades	0.003	0.010	0.053	0.253
Days Absent	11.742	12.583	14.685	34.807
Missing Attendance Data	0.000	0.000	0.002	0.176
Math Test Score (Z)	0.007	-0.098	-0.023	-0.034
Reading Test Score (Z)	0.059	-0.121	-0.019	-0.046
Missing Math Test	0.113	0.043	0.046	0.322
Missing Reading Test	0.117	0.045	0.052	0.323
Out-of-School Suspensions	0.188	0.369	0.829	4.718
Disciplinary Incidents	0.234	0.352	0.678	3.269
Any Arrests at Baseline	0.054	0.103	0.128	0.344
Arrests: Violent Crime	0.030	0.059	0.089	0.272
Arrests: Property Crime	0.006	0.030	0.069	0.158
Arrests: Drug Crime	0.010	0.018	0.025	0.158
Math Score - Decile in Previous School	5.926	5.484	5.807	5.915
In Study 2	0.430	0.525	0.536	0.388
Participated in Year 1 of Study	0.294	0.291	0.244	0.213

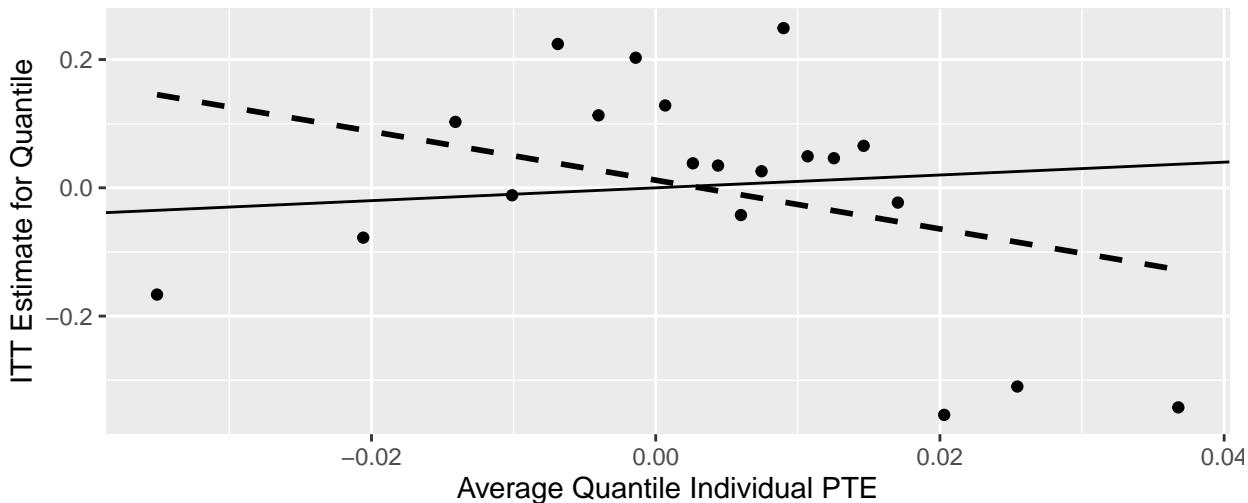
## Outcome: Reading test score (Z)



above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitors”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.003 (0.027).<sup>6</sup>

### Calibration Plot: readxil\_z\_post1\_np

PTE x Treatment interaction estimate is  $-4.324$  ( $1.742$ ), with a p-value of  $0.013$



<sup>6</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 0.993 and a ‘differential forest prediction’ (DFP) coefficient of -5.719, with a corresponding DFP p-value of 0.998. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 16: Average Treatment Effects (overall and for subsamples) for Reading test score (Z)

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.003 +/- 0.053	0.027
Individual PTE Quartile 4	-0.235 +/- 0.123	0.063
Individual PTE Quartile 3	0.113 +/- 0.106	0.054
Individual PTE Quartile 2	0.085 +/- 0.1	0.051
Individual PTE Quartile 1	0.043 +/- 0.096	0.049
Top 2 PTE Quartiles	-0.061 +/- 0.081	0.041
Bottom 2 PTE Quartiles	0.064 +/- 0.069	0.035
Bottom 3 PTE Quartiles	0.08 +/- 0.058	0.030

Table 17: Differences between subsample average treatment effects: Reading test score (Z)

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.452, -0.179]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.231, -0.018]
Quartile 4	Quartile 1	[-0.434, -0.122]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.660069887399302

Tuned parameters:

```
sample.fraction: 0.405681332328822
mtry: 1
honesty.fraction: 0.529774703970179
honesty.prune.leaves: 0
alpha: 0.0242021974408999
imbalance.penalty: 0.954647537496177
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0504,0.149] 0.6614364
(0.149,0.237] 0.6620166
(0.237,0.322] 0.6625534
(0.322,0.41] 0.6630067
(0.41,0.5] 0.6634239
```

```
mtry      error
[1,6] 0.6613729
(6,11] 0.6624242
(11,16] 0.6628859
(16,21] 0.6629431
(21,26] 0.6630992
```

```
honesty.fraction      error
[0.5,0.564] 0.6623007
(0.564,0.615] 0.6623766
(0.615,0.677] 0.6625114
(0.677,0.739] 0.6625816
```

```

(0.739,0.8] 0.6626667

honesty.prune.leaves      error
  0 0.6624662
  1 0.6625077

      alpha      error
[5.98e-05,0.0513] 0.6622133
(0.0513,0.0984] 0.6622858
(0.0984,0.147] 0.6623941
(0.147,0.201] 0.6626408
(0.201,0.25] 0.6629030

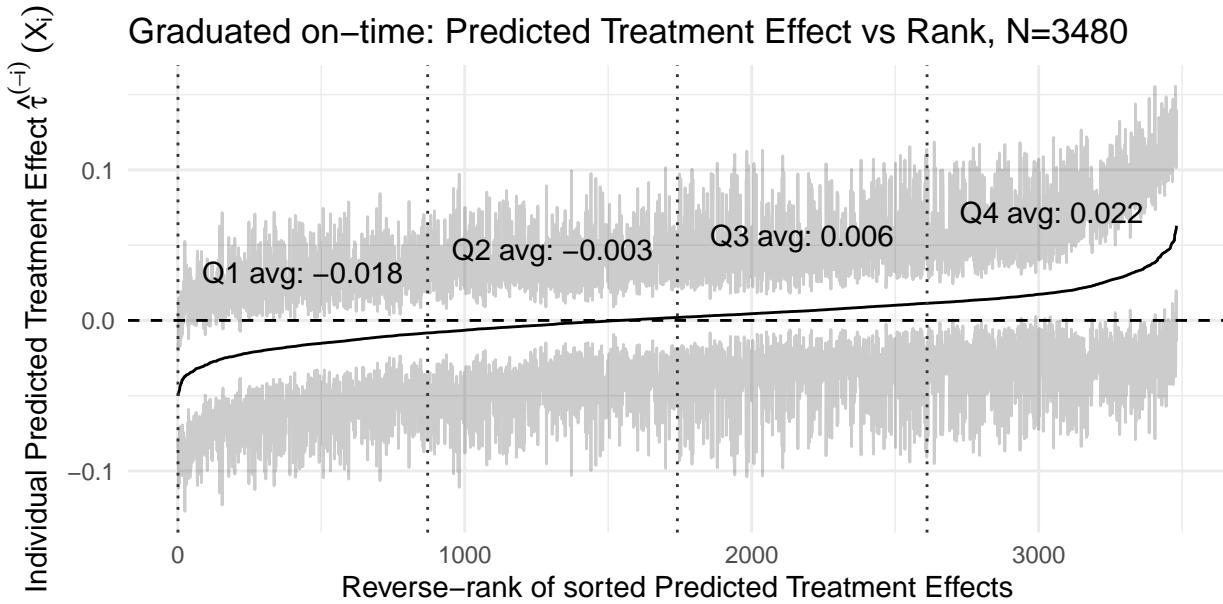
imbalance.penalty      error
[0.000325,0.229] 0.6625800
(0.229,0.513] 0.6625785
(0.513,0.934] 0.6626076
(0.934,1.6] 0.6623333
(1.6,7.71] 0.6623377

```

Table 18: Summary table by Quartile of Predicted Treatment Effects on Reading test score (Z)

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.017	0.000	0.009	0.023
N	899.000	898.000	898.000	899.000
Age	14.775	14.522	14.500	14.390
Female	0.080	0.143	0.220	0.195
Has IEP	0.492	0.136	0.033	0.017
Has Free/Reduced Lunch	0.960	0.937	0.895	0.778
Black	0.468	0.524	0.587	0.462
Hispanic	0.527	0.468	0.400	0.388
Other Race	0.004	0.008	0.013	0.150
In 9th Grade	0.642	0.714	0.755	0.803
In 10th Grade	0.353	0.286	0.245	0.196
Baseline GPA	1.945	2.278	2.567	2.831
Num. A's	4.246	6.051	8.208	10.913
Num. B's	5.473	7.336	8.694	8.360
Num. C's	9.089	9.253	8.657	5.960
Num. D's	5.263	3.680	2.536	1.593
Num. F's	3.667	2.555	1.865	1.319
Missing Baseline GPA/Grades	0.030	0.032	0.041	0.172
Days Absent	18.157	14.338	11.925	9.760
Missing Attendance Data	0.000	0.000	0.002	0.132
Math Test Score (Z)	-0.699	-0.202	0.289	0.874
Reading Test Score (Z)	-0.574	-0.108	0.315	0.616
Missing Math Test	0.016	0.022	0.046	0.246
Missing Reading Test	0.017	0.023	0.047	0.254
Out-of-School Suspensions	0.877	0.867	0.938	0.873
Disciplinary Incidents	0.887	0.685	0.642	0.622
Any Arrests at Baseline	0.150	0.128	0.078	0.057
Arrests: Violent Crime	0.057	0.079	0.051	0.050
Arrests: Property Crime	0.052	0.040	0.016	0.011
Arrests: Drug Crime	0.031	0.020	0.013	0.014
Math Score - Decile in Previous School	3.877	5.254	6.697	8.195
In Study 2	0.419	0.469	0.528	0.541
Participated in Year 1 of Study	0.308	0.255	0.252	0.253

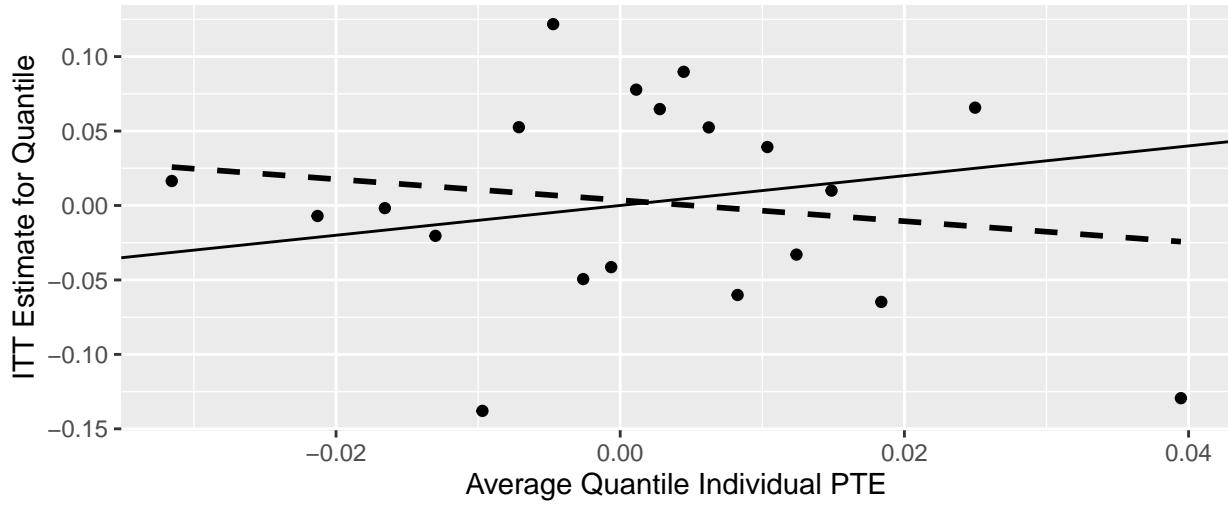
## Outcome: Graduated on-time



The above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.001 (0.012).<sup>7</sup>

### Calibration Plot: graduated\_ontime

PTE x Treatment interaction estimate is  $-0.848 (0.777)$ , with a p-value of 0.275



<sup>7</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.236 and a ‘differential forest prediction’ (DFP) coefficient of -1.09, with a corresponding DFP p-value of 0.898. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 19: Average Treatment Effects (overall and for subsamples) for Graduated on-time

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.001 +/- 0.024	0.012
Individual PTE Quartile 4	-0.053 +/- 0.051	0.026
Individual PTE Quartile 3	0.038 +/- 0.045	0.023
Individual PTE Quartile 2	0.043 +/- 0.049	0.025
Individual PTE Quartile 1	-0.023 +/- 0.051	0.026
Top 2 PTE Quartiles	-0.007 +/- 0.034	0.017
Bottom 2 PTE Quartiles	0.01 +/- 0.035	0.018
Bottom 3 PTE Quartiles	0.02 +/- 0.028	0.014

Table 20: Differences between subsample average treatment effects: Graduated on-time

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.131, -0.015]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.066, 0.032]
Quartile 4	Quartile 1	[-0.102, 0.042]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.13510507466622

Tuned parameters:

```
sample.fraction: 0.268119367142208
mtry: 5
honesty.fraction: 0.789731954806484
honesty.prune.leaves: 0
alpha: 0.0283485450199805
imbalance.penalty: 0.477790027150692
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0502,0.141] 0.1351471
(0.141,0.23] 0.1351093
(0.23,0.321] 0.1351234
(0.321,0.412] 0.1351690
(0.412,0.5] 0.1351970
```

```
mtry      error
[1,5] 0.1351057
(5,11] 0.1351382
(11,16] 0.1351361
(16,21] 0.1351976
(21,26] 0.1351789
```

```
honesty.fraction      error
[0.5,0.557] 0.1351436
(0.557,0.62] 0.1351511
(0.62,0.676] 0.1351532
(0.676,0.736] 0.1351586
```

```
(0.736,0.799] 0.1351392

honesty.prune.leaves      error
  0 0.1351474
  1 0.1351509

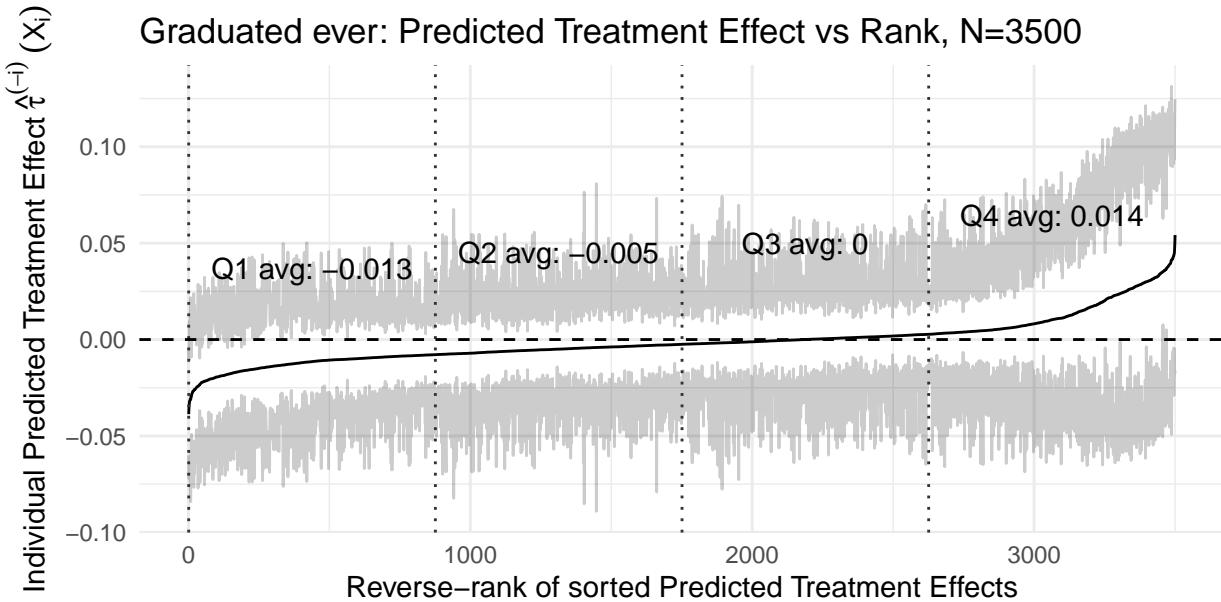
      alpha      error
[0.000195,0.0498] 0.1351347
(0.0498,0.102] 0.1351461
(0.102,0.153] 0.1351541
(0.153,0.202] 0.1351583
(0.202,0.25] 0.1351526

imbalance.penalty      error
[0.000405,0.23] 0.1351653
(0.23,0.484] 0.1351609
(0.484,0.882] 0.1351386
(0.882,1.58] 0.1351340
(1.58,5.66] 0.1351470
```

Table 21: Summary table by Quartile of Predicted Treatment Effects on Graduated on-time

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.018	-0.003	0.006	0.022
N	870.000	870.000	870.000	870.000
Age	14.932	14.532	14.351	14.530
Female	0.071	0.174	0.201	0.180
Has IEP	0.178	0.199	0.161	0.160
Has Free/Reduced Lunch	0.871	0.900	0.880	0.900
Black	0.509	0.517	0.515	0.620
Hispanic	0.447	0.445	0.437	0.351
Other Race	0.044	0.038	0.048	0.030
In 9th Grade	0.416	0.744	0.854	0.807
In 10th Grade	0.578	0.254	0.145	0.186
Baseline GPA	2.193	2.544	2.635	2.090
Num. A's	2.831	7.472	10.103	8.317
Num. B's	5.491	8.120	8.714	6.722
Num. C's	7.470	8.824	8.314	6.951
Num. D's	3.505	2.774	2.628	3.767
Num. F's	1.453	1.708	1.917	4.349
Missing Baseline GPA/Grades	0.025	0.080	0.098	0.069
Days Absent	13.713	12.244	11.281	24.041
Missing Attendance Data	0.001	0.021	0.051	0.052
Math Test Score (Z)	-0.045	-0.105	0.159	0.092
Reading Test Score (Z)	-0.021	-0.098	0.169	0.006
Missing Math Test	0.098	0.074	0.102	0.139
Missing Reading Test	0.102	0.075	0.108	0.145
Out-of-School Suspensions	0.063	0.170	0.419	4.053
Disciplinary Incidents	0.157	0.262	0.433	2.756
Any Arrests at Baseline	0.079	0.074	0.090	0.295
Arrests: Violent Crime	0.030	0.043	0.072	0.255
Arrests: Property Crime	0.029	0.032	0.068	0.139
Arrests: Drug Crime	0.015	0.017	0.036	0.098
Math Score - Decile in Previous School	5.738	5.536	6.203	6.220
In Study 2	0.330	0.499	0.575	0.520
Participated in Year 1 of Study	0.278	0.233	0.254	0.228

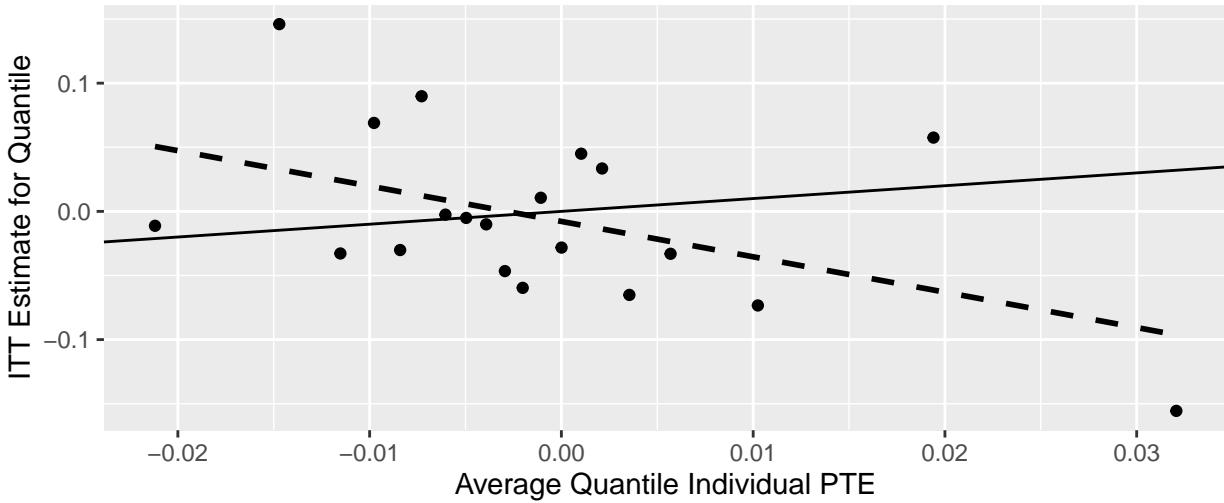
## Outcome: Graduated ever



above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of -0.001 (0.011).<sup>8</sup>

### Calibration Plot: graduated\_ever

PTE x Treatment interaction estimate is -3.109 (0.996), with a p-value of 0.002



<sup>8</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.286 and a ‘differential forest prediction’ (DFP) coefficient of -3.174, with a corresponding DFP p-value of 0.995. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 22: Average Treatment Effects (overall and for subsamples) for Graduated ever

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	-0.001 +/- 0.022	0.011
Individual PTE Quartile 4	-0.047 +/- 0.053	0.027
Individual PTE Quartile 3	0.009 +/- 0.039	0.020
Individual PTE Quartile 2	-0.004 +/- 0.037	0.019
Individual PTE Quartile 1	0.037 +/- 0.046	0.024
Top 2 PTE Quartiles	-0.019 +/- 0.033	0.017
Bottom 2 PTE Quartiles	0.016 +/- 0.03	0.015
Bottom 3 PTE Quartiles	0.014 +/- 0.024	0.012

Table 23: Differences between subsample average treatment effects: Graduated ever

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.118, -0.003]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.079, 0.009]
Quartile 4	Quartile 1	[-0.153, -0.013]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.109341750143883

Tuned parameters:

```
sample.fraction: 0.419261500006542
mtry: 3
honesty.fraction: 0.681272655189969
honesty.prune.leaves: 0
alpha: 0.0348542399005964
imbalance.penalty: 0.406441015093588
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0501,0.132] 0.1095809
(0.132,0.221] 0.1095563
(0.221,0.321] 0.1095651
(0.321,0.407] 0.1096106
(0.407,0.5] 0.1096403
```

```
mtry      error
[1,5] 0.1095300
(5,11] 0.1095599
(11,16] 0.1096133
(16,21] 0.1096248
(21,26] 0.1096390
```

```
honesty.fraction      error
[0.5,0.562] 0.1095820
(0.562,0.619] 0.1095972
(0.619,0.679] 0.1095823
(0.679,0.743] 0.1095903
```

```

(0.743,0.8] 0.1096015

honesty.prune.leaves      error
  0 0.1095905
  1 0.1095908

      alpha      error
[3.5e-06,0.0488] 0.1095646
(0.0488,0.0988] 0.1095827
(0.0988,0.15] 0.1095972
(0.15,0.197] 0.1096027
(0.197,0.25] 0.1096060

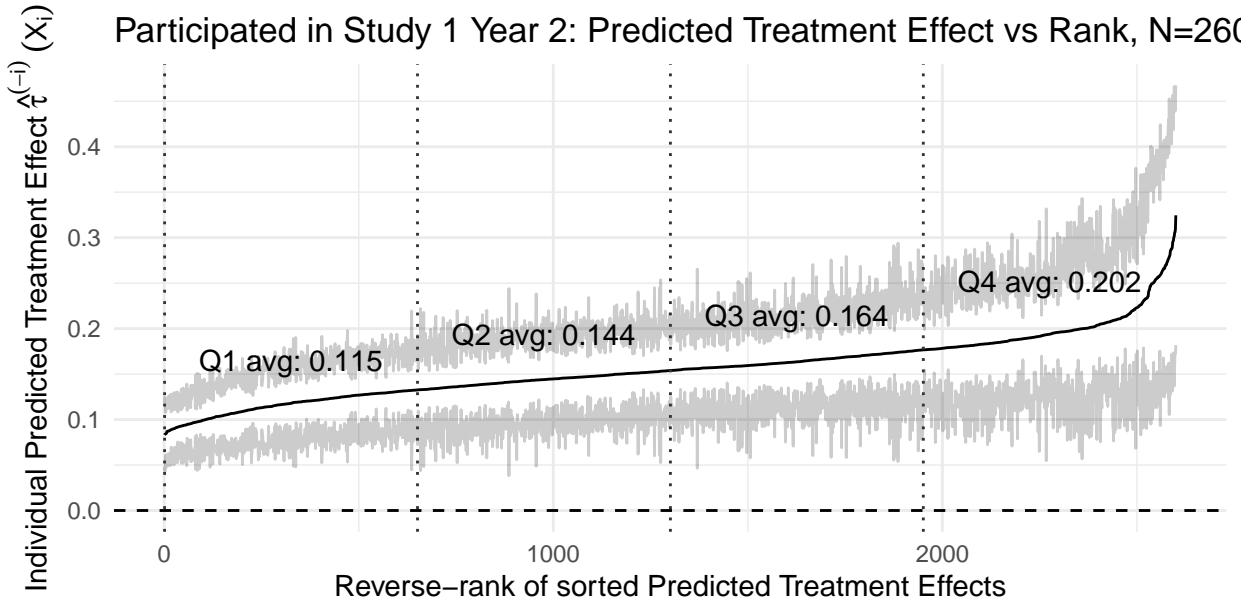
imbalance.penalty      error
[0.000439,0.223] 0.1096214
(0.223,0.476] 0.1095972
(0.476,0.854] 0.1095838
(0.854,1.51] 0.1095734
(1.51,6.96] 0.1095775

```

Table 24: Summary table by Quartile of Predicted Treatment Effects on Graduated ever

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.013	-0.005	0.000	0.014
N	875.000	875.000	875.000	875.000
Age	14.627	14.658	14.312	14.738
Female	0.120	0.157	0.186	0.162
Has IEP	0.162	0.178	0.158	0.199
Has Free/Reduced Lunch	0.874	0.873	0.904	0.898
Black	0.601	0.520	0.491	0.549
Hispanic	0.369	0.432	0.457	0.419
Other Race	0.030	0.048	0.051	0.032
In 9th Grade	0.709	0.594	0.821	0.682
In 10th Grade	0.283	0.401	0.173	0.299
Baseline GPA	2.180	2.548	2.726	1.936
Num. A's	4.291	6.917	11.365	6.142
Num. B's	6.586	6.987	9.855	5.560
Num. C's	10.057	7.530	7.942	6.031
Num. D's	4.366	2.806	2.400	3.141
Num. F's	2.622	1.577	1.467	3.847
Missing Baseline GPA/Grades	0.018	0.053	0.045	0.158
Days Absent	12.327	11.666	11.078	27.630
Missing Attendance Data	0.000	0.000	0.000	0.123
Math Test Score (Z)	0.126	0.179	0.167	-0.554
Reading Test Score (Z)	-0.014	0.149	0.196	-0.395
Missing Math Test	0.010	0.039	0.058	0.305
Missing Reading Test	0.011	0.042	0.064	0.312
Out-of-School Suspensions	0.288	0.335	0.393	3.966
Disciplinary Incidents	0.224	0.286	0.347	2.949
Any Arrests at Baseline	0.111	0.074	0.064	0.293
Arrests: Violent Crime	0.058	0.037	0.043	0.271
Arrests: Property Crime	0.032	0.022	0.040	0.174
Arrests: Drug Crime	0.016	0.016	0.026	0.117
Math Score - Decile in Previous School	6.333	6.342	6.161	4.293
In Study 2	0.470	0.447	0.547	0.464
Participated in Year 1 of Study	0.256	0.230	0.247	0.255

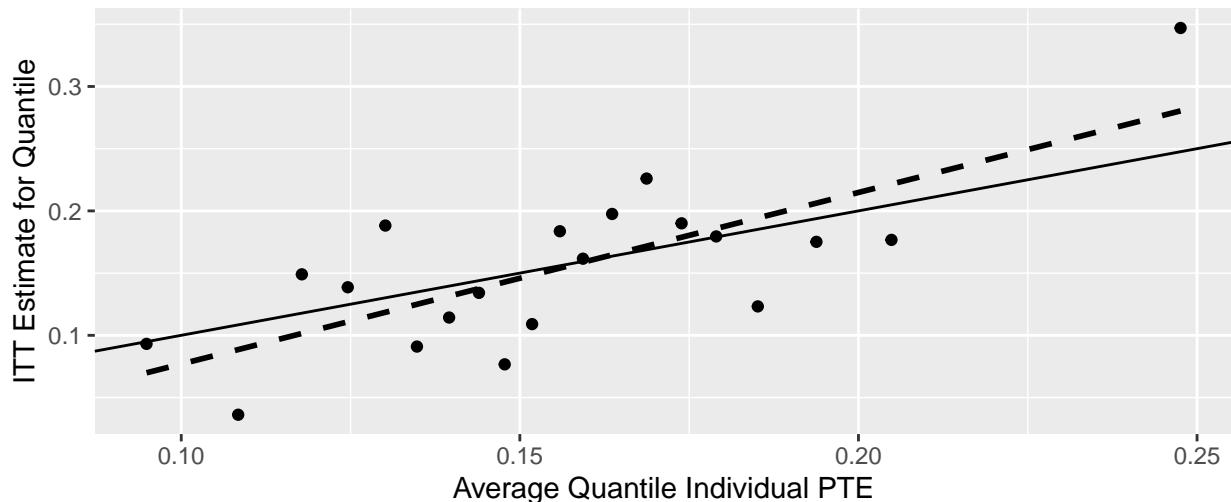
## Outcome: Participated in Study 1 Year 2



The above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitors”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ 's  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.156

### Calibration Plot: treat\_post2

PTE x Treatment interaction estimate is 1.439 (0.288), with a p-value of 0



(0.01).<sup>9</sup>

<sup>9</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 0.997 and a ‘differential forest prediction’ (DFP) coefficient of 1.376, with a corresponding DFP p-value of 0. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 25: Average Treatment Effects (overall and for subsamples) for Participated in Study 1 Year 2

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.156 +/- 0.02	0.010
Individual PTE Quartile 4	0.198 +/- 0.044	0.023
Individual PTE Quartile 3	0.196 +/- 0.044	0.023
Individual PTE Quartile 2	0.105 +/- 0.036	0.018
Individual PTE Quartile 1	0.124 +/- 0.036	0.018
Top 2 PTE Quartiles	0.197 +/- 0.031	0.016
Bottom 2 PTE Quartiles	0.114 +/- 0.025	0.013
Bottom 3 PTE Quartiles	0.142 +/- 0.023	0.011

Table 26: Differences between subsample average treatment effects: Participated in Study 1 Year 2

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[0.006, 0.106]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[0.043, 0.124]
Quartile 4	Quartile 1	[0.017, 0.131]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.0686387474075786

Tuned parameters:

```
sample.fraction: 0.455612582969479
mtry: 8
honesty.fraction: 0.738270075223409
honesty.prune.leaves: 0
alpha: 0.0148619895335287
imbalance.penalty: 0.0459479139243858
```

Average error by 5-quantile:

```
sample.fraction      error
[0.05,0.135] 0.06914198
(0.135,0.223] 0.06908791
(0.223,0.307] 0.06907231
(0.307,0.403] 0.06905041
(0.403,0.5] 0.06901132
```

```
mtry      error
[1,6] 0.06911168
(6,11] 0.06906338
(11,16] 0.06906829
(16,21] 0.06906496
(21,26] 0.06905073
```

```
honesty.fraction      error
[0.5,0.557] 0.06909999
(0.557,0.613] 0.06907965
(0.613,0.676] 0.06907089
(0.676,0.738] 0.06905593
```

```

(0.738,0.8] 0.06905747

honesty.prune.leaves      error
  0 0.0690703
  1 0.0690751

      alpha      error
[9.3e-06,0.0577] 0.06899800
(0.0577,0.104] 0.06906858
(0.104,0.153] 0.06909223
(0.153,0.202] 0.06909790
(0.202,0.25] 0.06910723

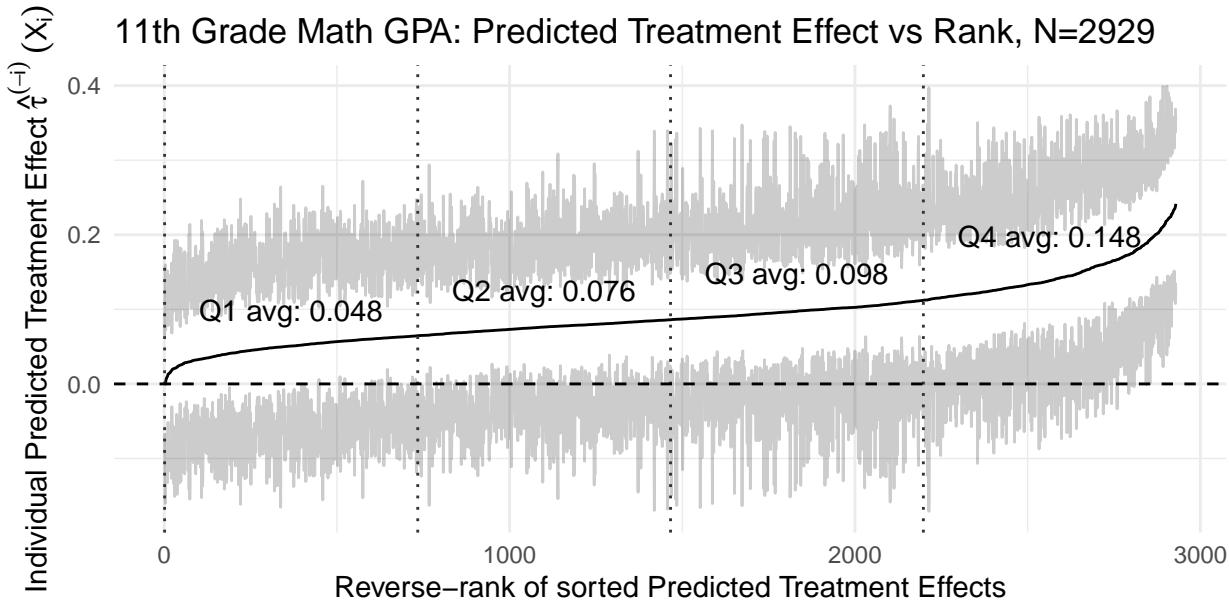
imbalance.penalty      error
[0.0012,0.233] 0.06900555
(0.233,0.512] 0.06902338
(0.512,0.908] 0.06906098
(0.908,1.58] 0.06910904
(1.58,7.69] 0.06916498

```

Table 27: Summary table by Quartile of Predicted Treatment Effects on Participated in Study 1 Year 2

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.115	0.144	0.164	0.202
N	650.000	650.000	650.000	650.000
Age	14.986	14.786	14.758	14.710
Female	0.002	0.000	0.000	0.003
Has IEP	0.068	0.072	0.180	0.365
Has Free/Reduced Lunch	0.857	0.894	0.886	0.825
Black	0.392	0.455	0.532	0.482
Hispanic	0.606	0.532	0.445	0.348
Other Race	0.002	0.012	0.023	0.171
In 9th Grade	0.532	0.546	0.546	0.620
In 10th Grade	0.443	0.438	0.451	0.375
Baseline GPA	2.009	2.020	2.161	2.214
Num. A's	7.017	4.565	4.695	4.620
Num. B's	5.325	5.565	5.738	6.203
Num. C's	2.138	6.388	8.234	9.217
Num. D's	1.583	3.662	4.049	3.945
Num. F's	4.086	3.145	2.549	1.775
Missing Baseline GPA/Grades	0.080	0.058	0.042	0.111
Days Absent	30.089	19.228	17.392	16.060
Missing Attendance Data	0.000	0.000	0.020	0.105
Math Test Score (Z)	0.714	0.278	-0.164	-0.903
Reading Test Score (Z)	0.584	0.151	-0.277	-0.423
Missing Math Test	0.197	0.125	0.114	0.186
Missing Reading Test	0.200	0.126	0.118	0.183
Out-of-School Suspensions	2.357	1.488	1.688	1.010
Disciplinary Incidents	1.726	1.314	1.184	0.895
Any Arrests at Baseline	0.288	0.180	0.160	0.106
Arrests: Violent Crime	0.238	0.137	0.091	0.075
Arrests: Property Crime	0.154	0.080	0.062	0.042
Arrests: Drug Crime	0.154	0.057	0.035	0.028
Math Score - Decile in Previous School	8.252	7.366	5.713	3.154
In Study 2	0.000	0.000	0.000	0.000
Participated in Year 1 of Study	0.128	0.195	0.262	0.228

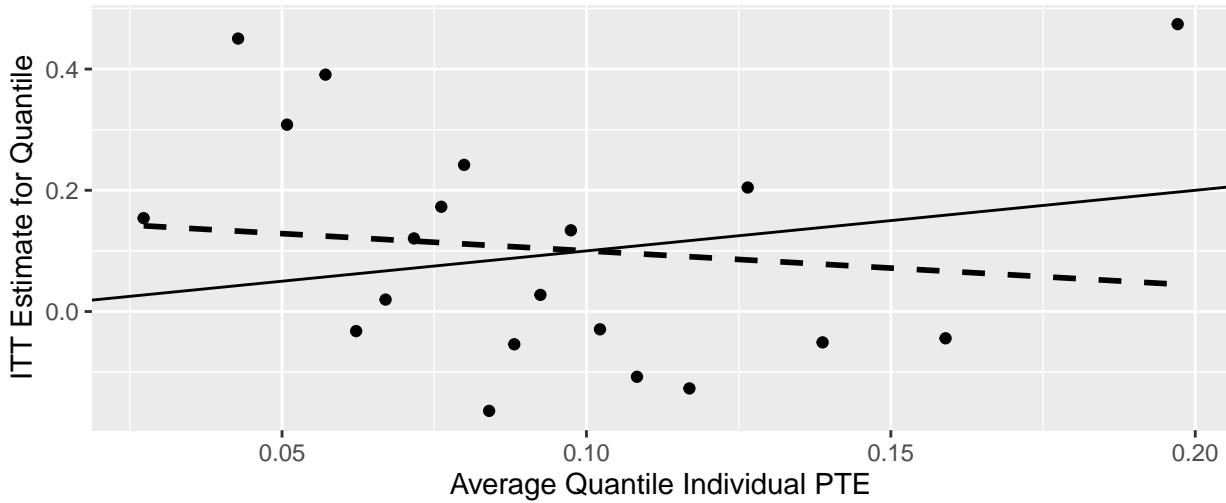
## Outcome: 11th Grade Math GPA



The above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitters”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.101 (0.038).<sup>10</sup>

### Calibration Plot: gpa11\_math

PTE x Treatment interaction estimate is  $-0.417$  (0.935), with a p-value of 0.656



<sup>10</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.048 and a ‘differential forest prediction’ (DFP) coefficient of -0.526, with a corresponding DFP p-value of 0.71. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 28: Average Treatment Effects (overall and for subsamples) for 11th Grade Math GPA

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.101 +/- 0.074	0.038
Individual PTE Quartile 4	0.111 +/- 0.147	0.075
Individual PTE Quartile 3	-0.058 +/- 0.154	0.079
Individual PTE Quartile 2	0.06 +/- 0.149	0.076
Individual PTE Quartile 1	0.286 +/- 0.144	0.074
Top 2 PTE Quartiles	0.027 +/- 0.107	0.054
Bottom 2 PTE Quartiles	0.173 +/- 0.103	0.053
Bottom 3 PTE Quartiles	0.098 +/- 0.086	0.044

Table 29: Differences between subsample average treatment effects: 11th Grade Math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.157, 0.184]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.295, 0.002]
Quartile 4	Quartile 1	[-0.381, 0.031]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 1.04737688336595

Tuned parameters:

```
sample.fraction: 0.154913748882245
mtry: 4
honesty.fraction: 0.727229274646379
honesty.prune.leaves: 0
alpha: 0.0800580532522872
imbalance.penalty: 1.13349636939617
```

Average error by 5-quantile:

```
sample.fraction      error
[0.0501,0.145]  1.047447
(0.145,0.228]   1.047833
(0.228,0.324]   1.048460
(0.324,0.411]   1.049037
(0.411,0.5]     1.049170
```

```
mtry      error
[1,6]  1.047734
(6,11] 1.048425
(11,16] 1.048890
(16,21] 1.048498
(21,26] 1.048450
```

```
honesty.fraction      error
[0.5,0.558]  1.048243
(0.558,0.618] 1.048352
(0.618,0.673] 1.048398
(0.673,0.735] 1.048434
```

```

(0.735,0.8] 1.048519

honesty.prune.leaves      error
  0 1.048364
  1 1.048415

      alpha      error
[0.000109,0.0484] 1.048337
(0.0484,0.0983] 1.048332
(0.0983,0.146] 1.048426
(0.146,0.198] 1.048398
(0.198,0.25] 1.048453

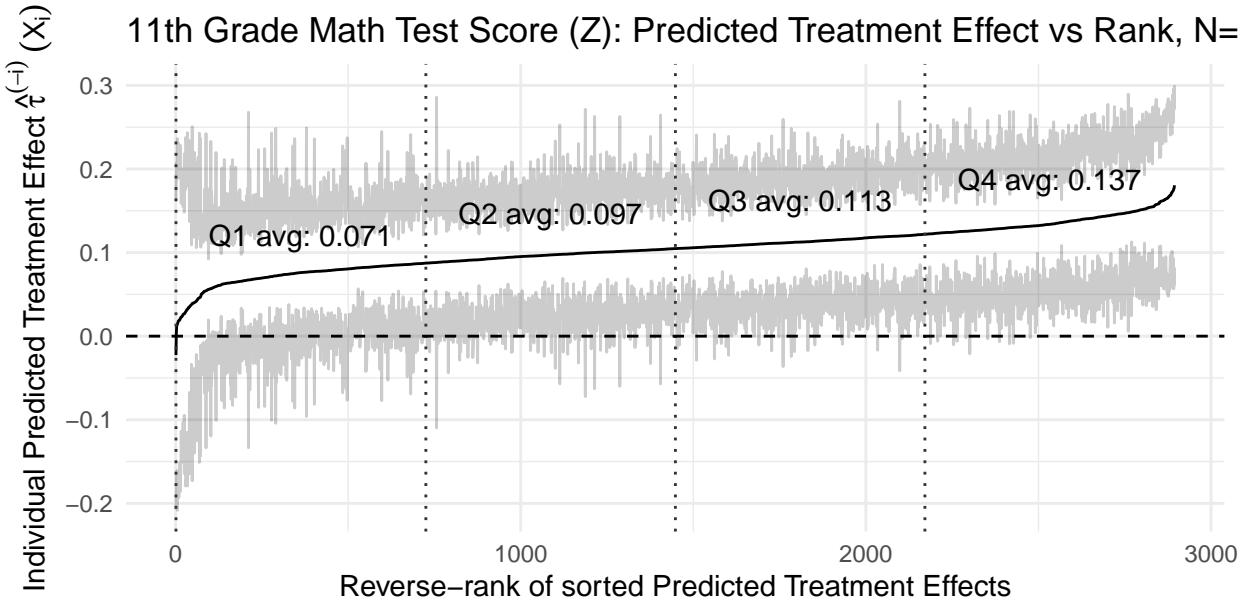
imbalance.penalty      error
[9.78e-05,0.212] 1.048386
(0.212,0.511] 1.048359
(0.511,0.928] 1.048395
(0.928,1.62] 1.048373
(1.62,10.2] 1.048433

```

Table 30: Summary table by Quartile of Predicted Treatment Effects on 11th Grade Math GPA

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.048	0.076	0.098	0.148
N	733.000	732.000	732.000	732.000
Age	14.198	14.331	14.564	15.124
Female	0.186	0.186	0.163	0.071
Has IEP	0.108	0.152	0.210	0.246
Has Free/Reduced Lunch	0.881	0.883	0.892	0.885
Black	0.765	0.503	0.407	0.303
Hispanic	0.199	0.451	0.534	0.645
Other Race	0.035	0.046	0.059	0.052
In 9th Grade	0.969	0.876	0.713	0.183
In 10th Grade	0.030	0.123	0.277	0.814
Baseline GPA	2.360	2.610	2.519	2.428
Num. A's	8.262	10.537	8.327	3.451
Num. B's	9.662	9.407	7.556	4.051
Num. C's	11.397	8.579	7.346	5.020
Num. D's	3.789	2.865	3.045	2.460
Num. F's	2.843	2.251	1.825	0.682
Missing Baseline GPA/Grades	0.031	0.042	0.087	0.094
Days Absent	14.197	11.881	12.153	11.352
Missing Attendance Data	0.000	0.007	0.060	0.064
Math Test Score (Z)	0.363	0.136	-0.057	-0.273
Reading Test Score (Z)	0.332	0.115	-0.114	-0.209
Missing Math Test	0.065	0.068	0.109	0.105
Missing Reading Test	0.071	0.070	0.112	0.108
Out-of-School Suspensions	1.142	0.659	0.679	0.390
Disciplinary Incidents	0.668	0.634	0.734	0.559
Any Arrests at Baseline	0.116	0.074	0.100	0.075
Arrests: Violent Crime	0.070	0.053	0.059	0.033
Arrests: Property Crime	0.037	0.031	0.033	0.026
Arrests: Drug Crime	0.026	0.014	0.023	0.018
Math Score - Decile in Previous School	6.934	6.140	5.751	5.051
In Study 2	0.588	0.564	0.508	0.225
Participated in Year 1 of Study	0.262	0.247	0.316	0.280

## Outcome: 11th Grade Math Test Score (Z)

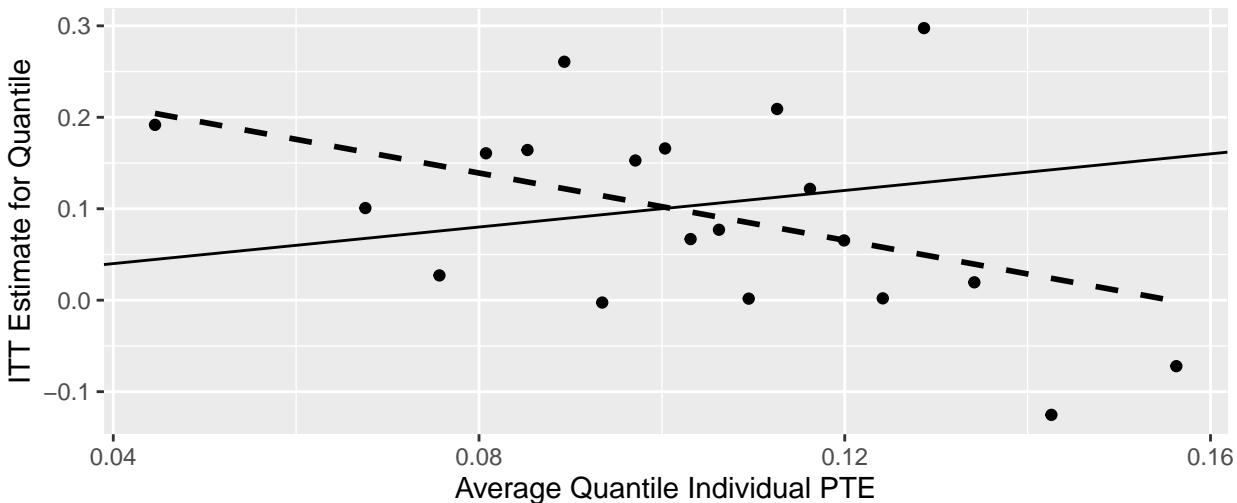


The

above plot shows the individual predicted treatment effects (PTEs) in rank-order from smallest to largest. The quartiles are labeled in the above plot, with Quartile 4 being the predicted “largest benefitors”, and Quartile 1 being those expected to see the smallest treatment effects. PTEs are also denoted  $\hat{\tau}^{(-i)}(X_i)$ , or the estimate of individual  $i$ ’s  $\hat{\tau}$  calculated from our fitted forest. The superscript  $(-i)$  denotes cross-fitting, eg ‘that the observation is computed by leaving observation  $i$  out’. We also can calculate the average treatment effect (ATE)  $\hat{\tau}$  from our fitted forest (done using the double robust method presented by Athey et. al. We present these ATEs below, including the ATEs calculated for those students identified in our PTE quartiles. Our causal forest estimates an overall average treatment effect of 0.104 (0.028).<sup>11</sup>

### Calibration Plot: eleventh\_grade\_math\_z

PTE x Treatment interaction estimate is -2.312 (1.134), with a p-value of 0.041



<sup>11</sup>We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.009 and a ‘differential forest prediction’ (DFP) coefficient of -3.508, with a corresponding DFP p-value of 0.998. A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 ‘additionally suggests that the forest has captured heterogeneity in the underlying signal.’ The p-value from the DFP estimate ‘acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity’.

Table 31: Average Treatment Effects (overall and for subsamples) for 11th Grade Math Test Score (Z)

Sample	Avg. Treatment Effect with 95% Conf. Intervals	Standard Error
Whole Sample	0.104 +/- 0.055	0.028
Individual PTE Quartile 4	0.035 +/- 0.113	0.058
Individual PTE Quartile 3	0.067 +/- 0.11	0.056
Individual PTE Quartile 2	0.131 +/- 0.112	0.057
Individual PTE Quartile 1	0.188 +/- 0.105	0.054
Top 2 PTE Quartiles	0.05 +/- 0.079	0.040
Bottom 2 PTE Quartiles	0.159 +/- 0.077	0.039
Bottom 3 PTE Quartiles	0.129 +/- 0.063	0.032

Table 32: Differences between subsample average treatment effects: 11th Grade Math Test Score (Z)

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.223, 0.036]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.219, 0.001]
Quartile 4	Quartile 1	[-0.307, 0.002]

Tuning status: tuned.

This indicates tuning found parameters that are expected to perform better than default.

Predicted debiased error: 0.56100746267978

Tuned parameters:

```
sample.fraction: 0.359763981180731
mtry: 2
honesty.fraction: 0.721831896109506
honesty.prune.leaves: 0
alpha: 0.00310937518952414
imbalance.penalty: 0.273028826623677
```

Average error by 5-quantile:

```
sample.fraction      error
[0.05,0.138] 0.5637958
(0.138,0.233] 0.5637889
(0.233,0.32] 0.5640983
(0.32,0.409] 0.5645980
(0.409,0.5] 0.5651320
```

```
mtry      error
[1,5] 0.5632573
(5,11] 0.5638442
(11,16] 0.5645761
(16,21] 0.5649778
(21,26] 0.5649485
```

```
honesty.fraction      error
[0.5,0.558] 0.5641244
(0.558,0.622] 0.5641465
(0.622,0.682] 0.5643120
(0.682,0.738] 0.5643528
```

```

(0.738,0.8] 0.5644772

honesty.prune.leaves      error
  0 0.5642724
  1 0.5642932

      alpha      error
[0.00011,0.0447] 0.5641363
(0.0447,0.0968] 0.5642027
(0.0968,0.152] 0.5643138
(0.152,0.202] 0.5643329
(0.202,0.25] 0.5644273

imbalance.penalty      error
[0.00283,0.252] 0.5643730
(0.252,0.548] 0.5644097
(0.548,0.935] 0.5643660
(0.935,1.61] 0.5641631
(1.61,8.24] 0.5641011

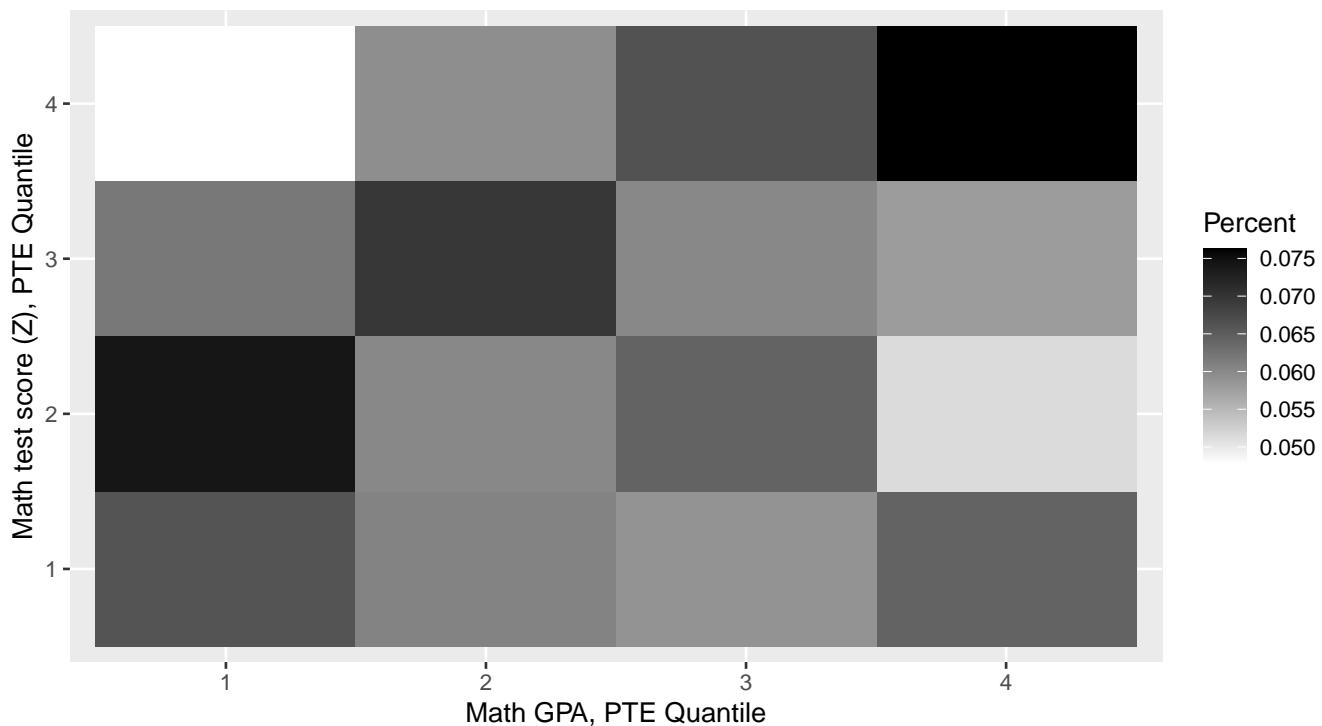
```

Table 33: Summary table by Quartile of Predicted Treatment Effects on 11th Grade Math Test Score (Z)

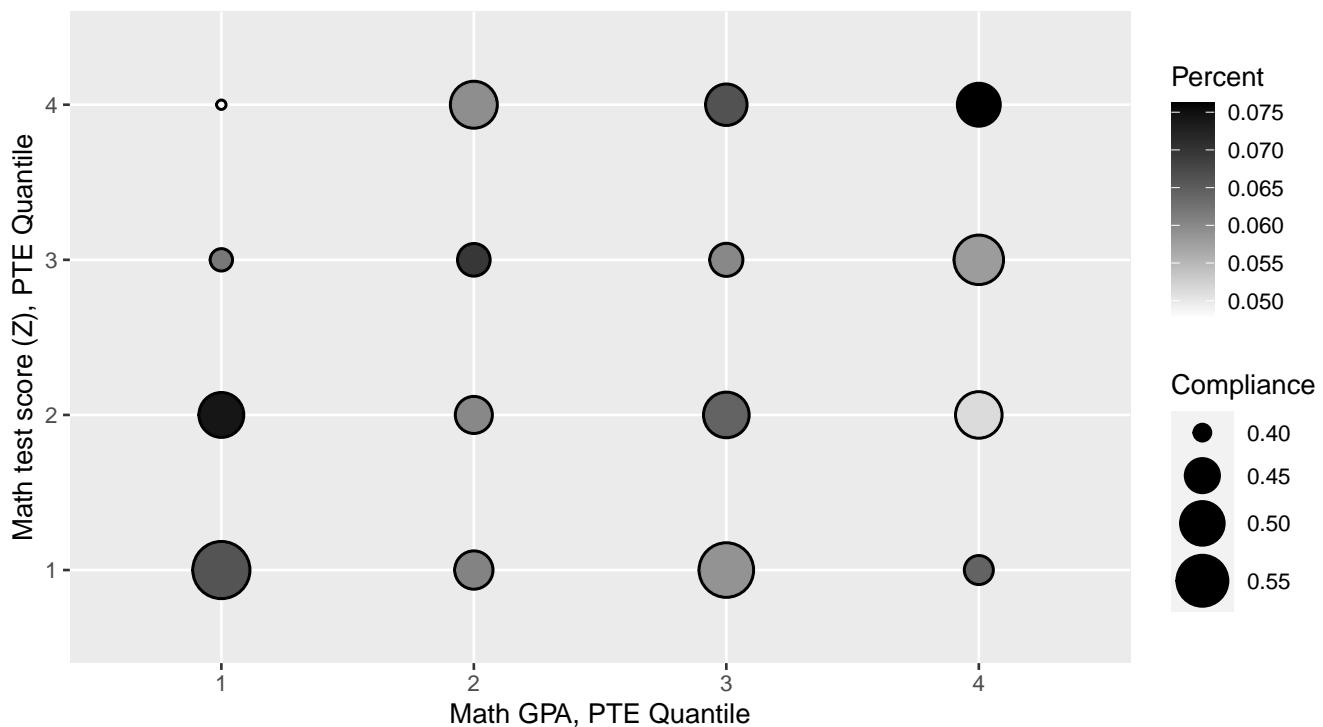
Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.071	0.097	0.113	0.137
N	724.000	723.000	723.000	724.000
Age	14.590	14.390	14.408	14.704
Female	0.229	0.194	0.134	0.032
Has IEP	0.316	0.214	0.116	0.052
Has Free/Reduced Lunch	0.902	0.877	0.876	0.887
Black	0.678	0.566	0.497	0.242
Hispanic	0.302	0.409	0.461	0.664
Other Race	0.019	0.025	0.043	0.094
In 9th Grade	0.834	0.844	0.732	0.377
In 10th Grade	0.166	0.156	0.268	0.623
Baseline GPA	2.091	2.487	2.745	2.678
Num. A's	5.039	8.235	10.456	7.881
Num. B's	7.536	8.972	8.895	6.319
Num. C's	8.559	9.598	8.364	6.186
Num. D's	4.030	3.189	2.275	2.275
Num. F's	2.953	2.008	1.149	0.950
Missing Baseline GPA/Grades	0.162	0.054	0.024	0.017
Days Absent	17.619	11.798	9.185	8.306
Missing Attendance Data	0.115	0.004	0.000	0.000
Math Test Score (Z)	-0.574	-0.132	0.320	0.607
Reading Test Score (Z)	-0.481	-0.058	0.219	0.484
Missing Math Test	0.185	0.040	0.019	0.025
Missing Reading Test	0.185	0.044	0.025	0.025
Out-of-School Suspensions	2.125	0.506	0.163	0.046
Disciplinary Incidents	1.671	0.492	0.248	0.075
Any Arrests at Baseline	0.159	0.071	0.057	0.039
Arrests: Violent Crime	0.108	0.041	0.017	0.006
Arrests: Property Crime	0.051	0.018	0.018	0.015
Arrests: Drug Crime	0.025	0.011	0.007	0.004
Math Score - Decile in Previous School	4.136	5.341	6.773	7.768
In Study 2	0.550	0.552	0.508	0.283
Participated in Year 1 of Study	0.286	0.257	0.248	0.262

## Decile Heatmaps

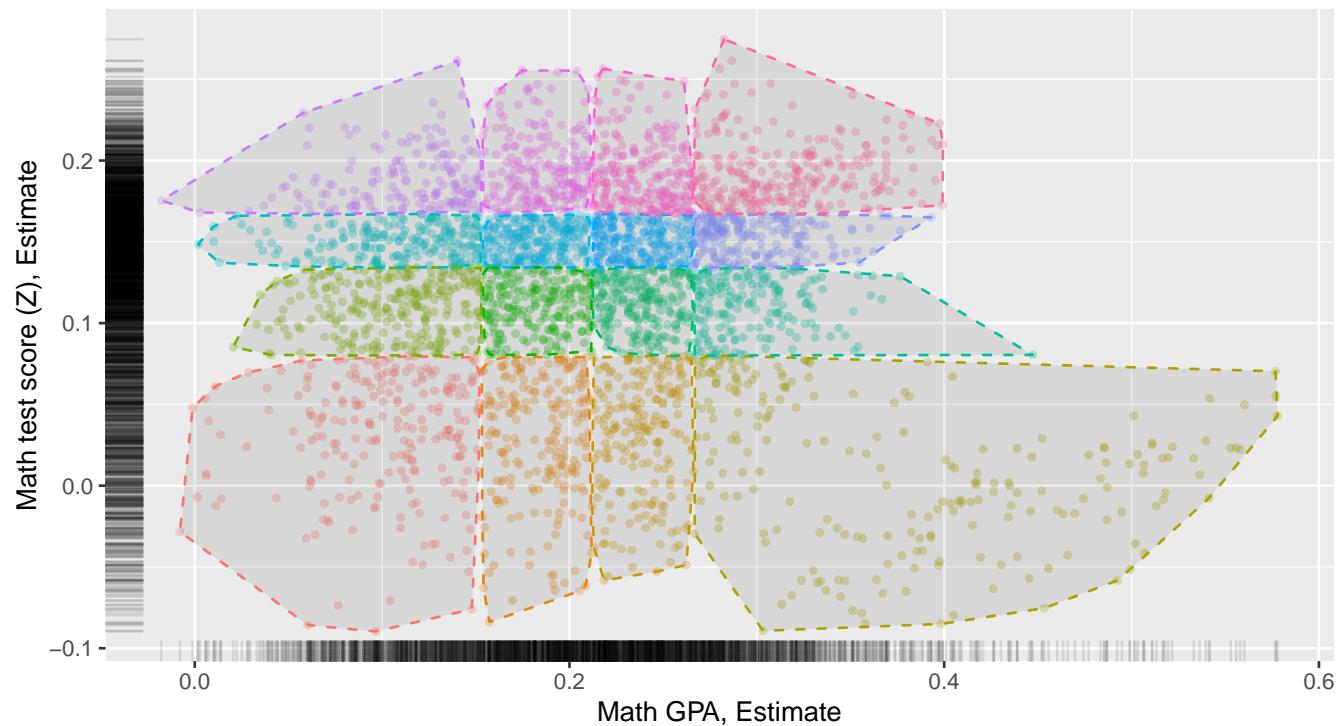
PTE Decile Plot: Math test score (Z) vs. Math GPA



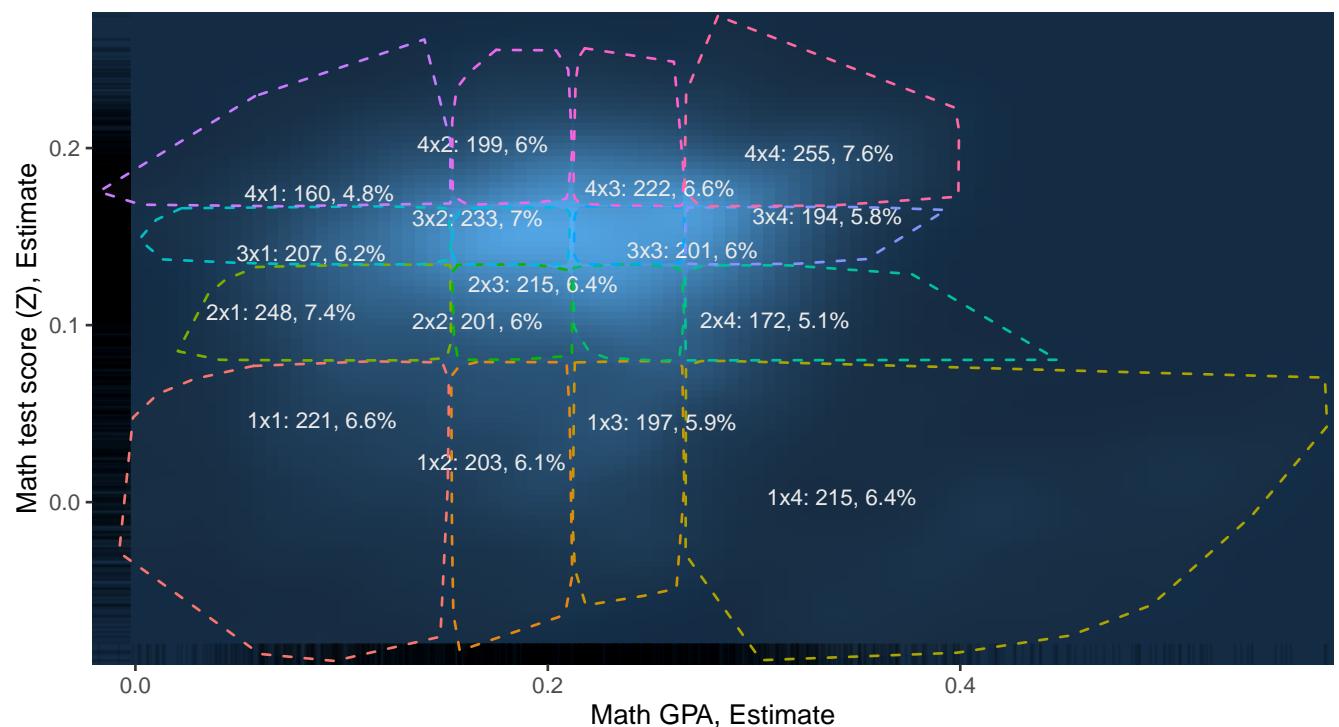
PTE Decile Plot: Math test score (Z) vs. Math GPA



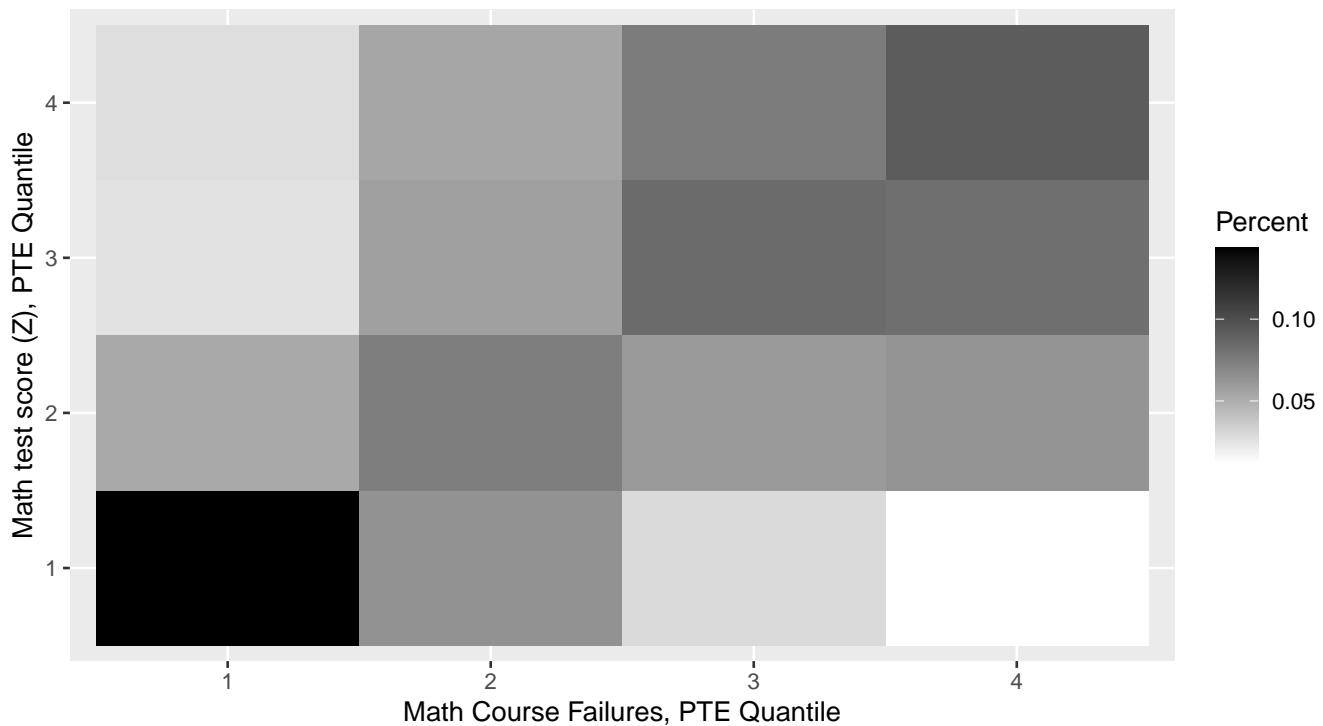
PTE Scatterplot: Math test score (Z) vs Math GPA



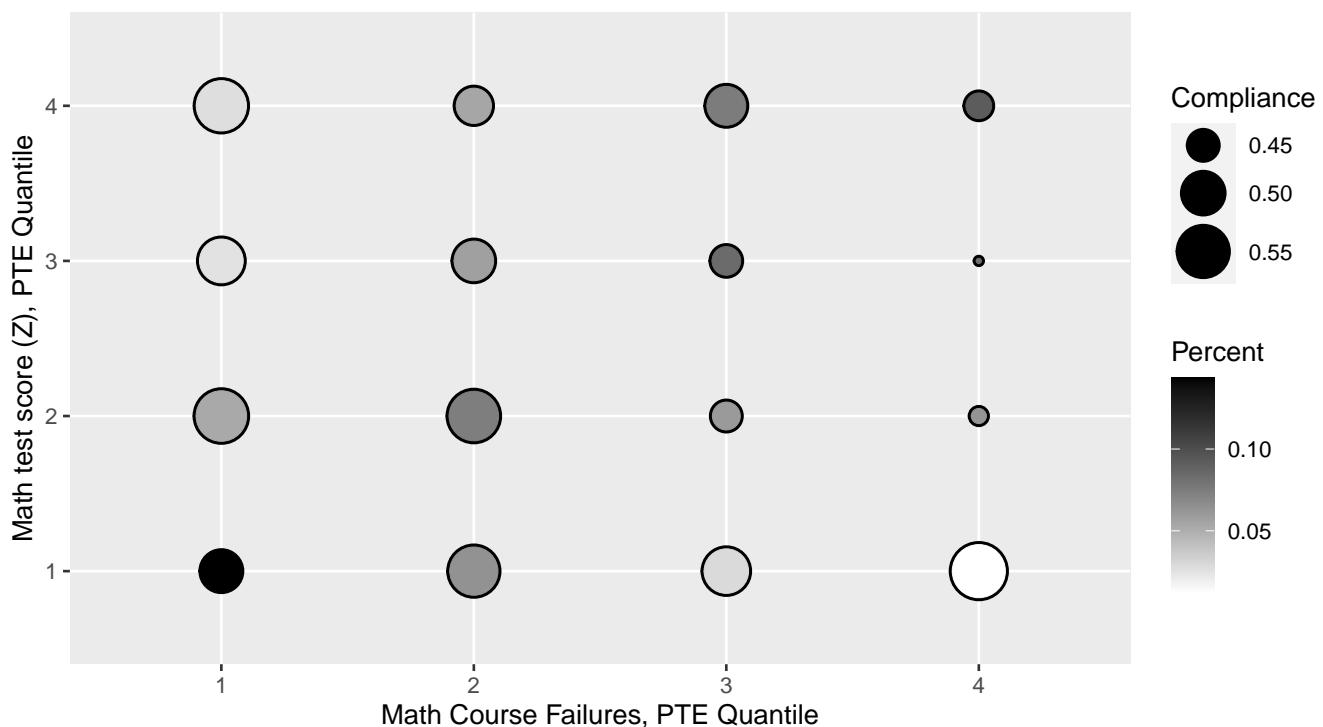
PTE Densities: Math test score (Z) vs Math GPA



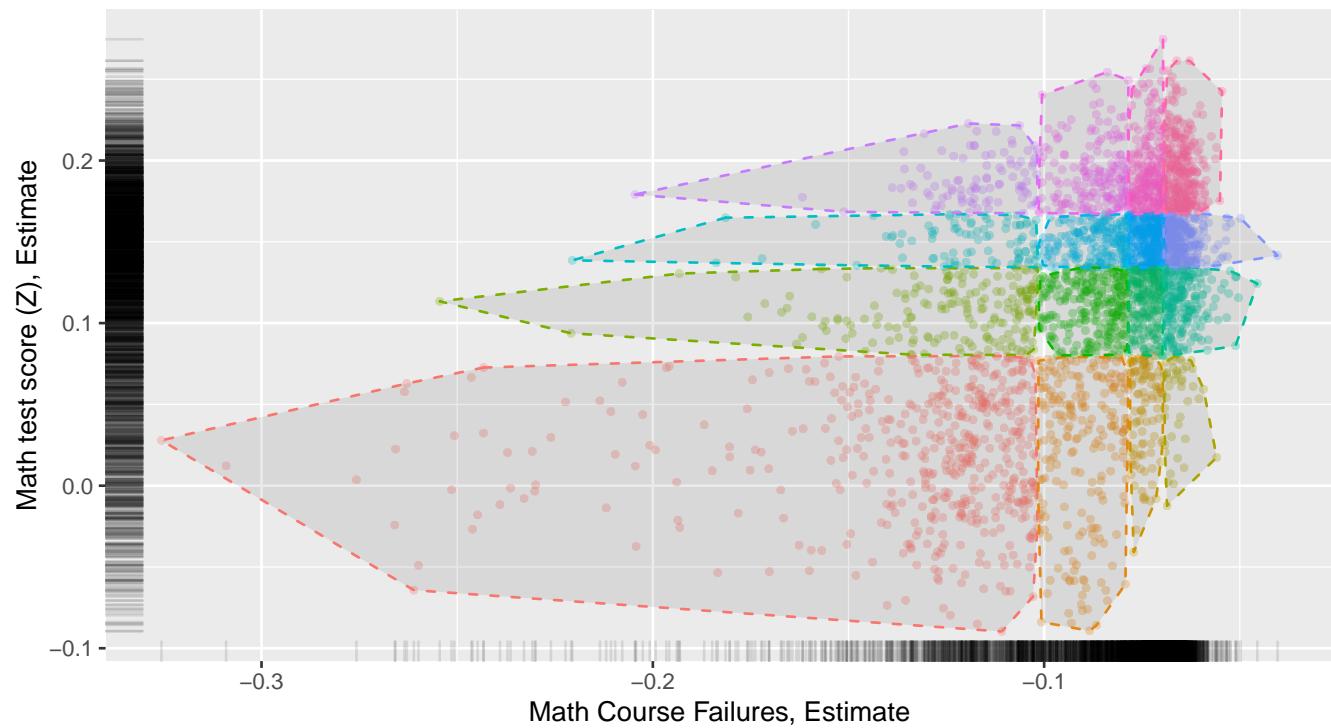
PTE Decile Plot: Math test score (Z) vs. Math Course Failures



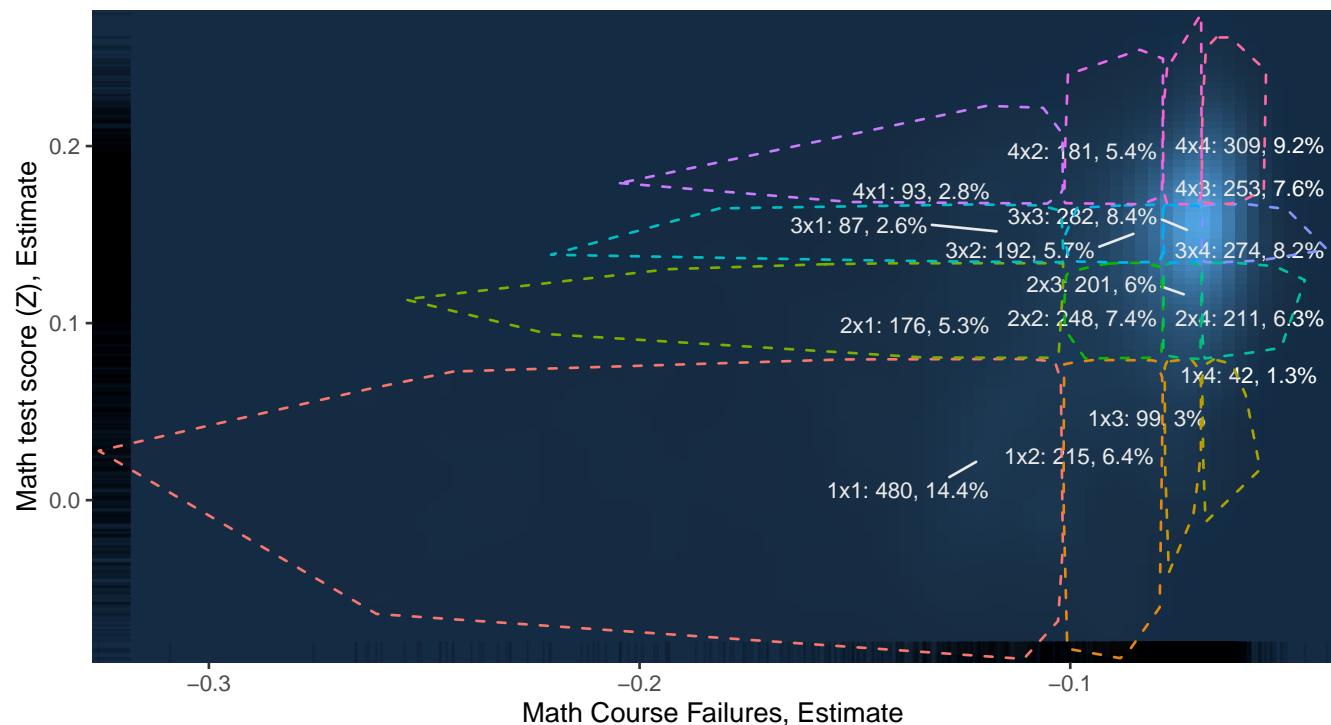
PTE Decile Plot: Math test score (Z) vs. Math Course Failures



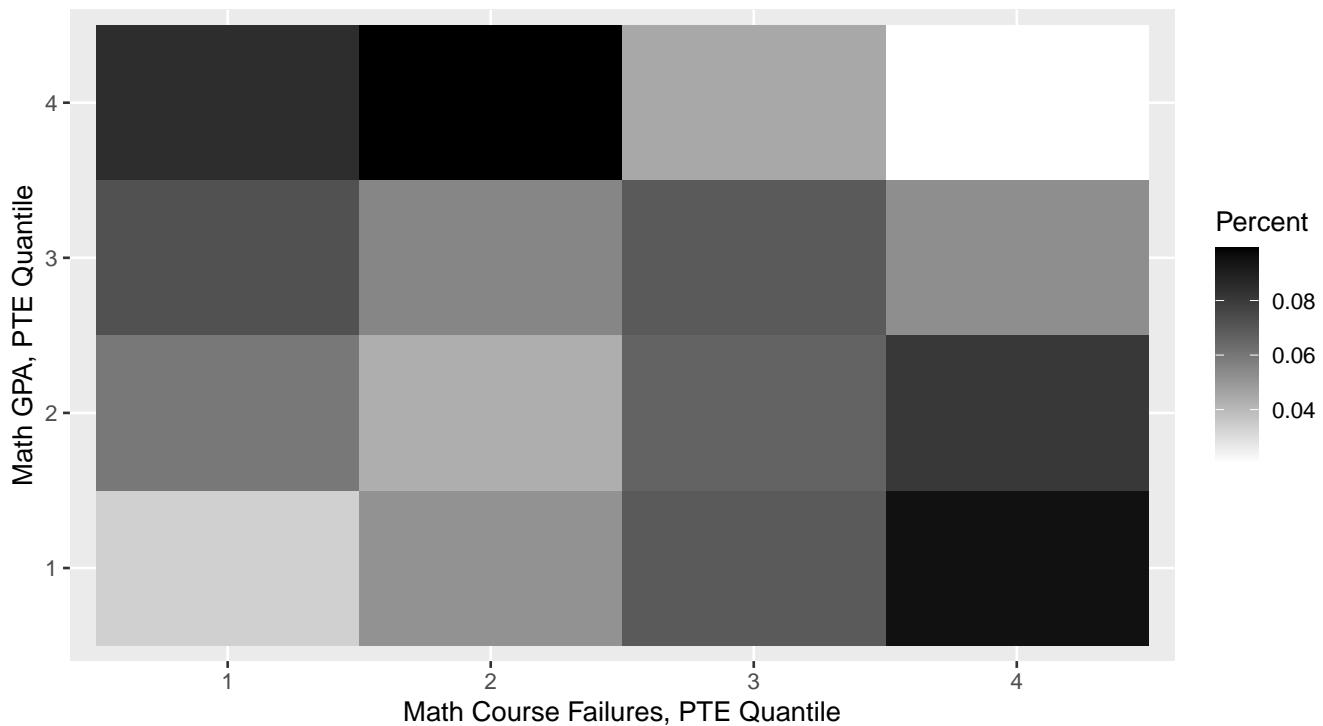
PTE Scatterplot: Math test score (Z) vs Math Course Failures



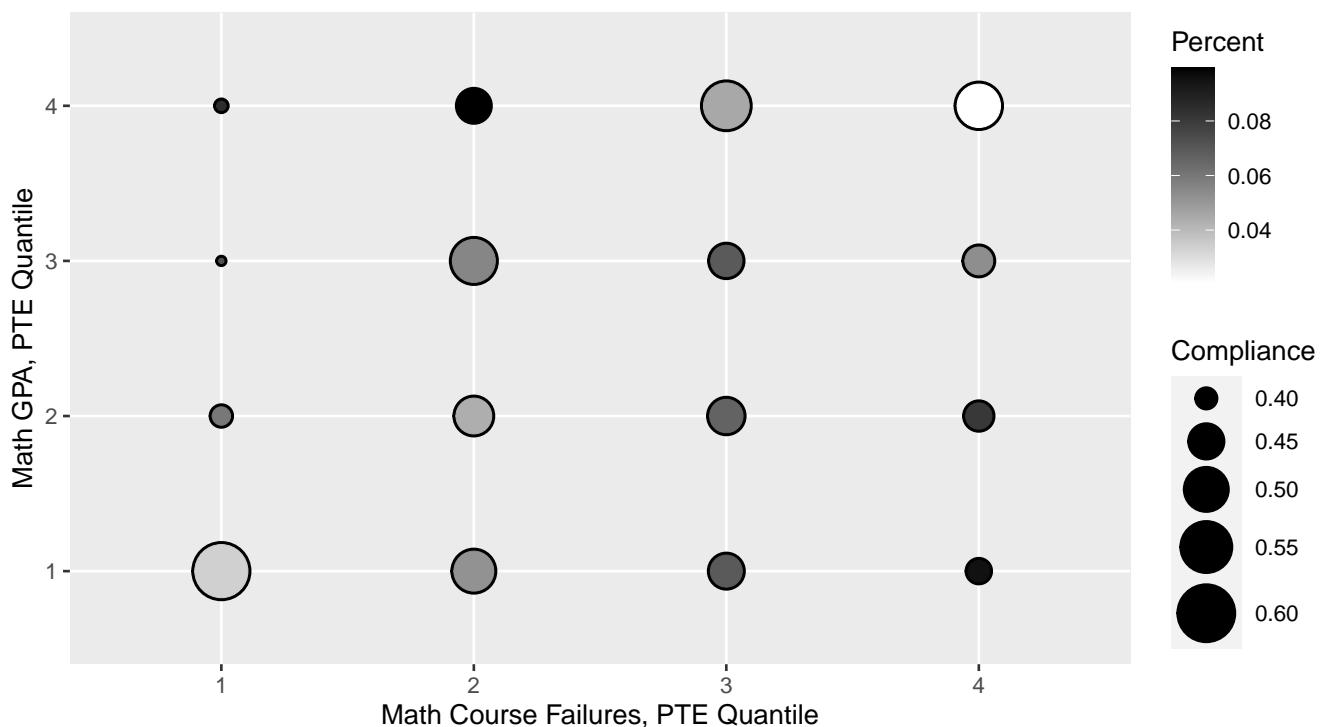
PTE Densities: Math test score (Z) vs Math Course Failures



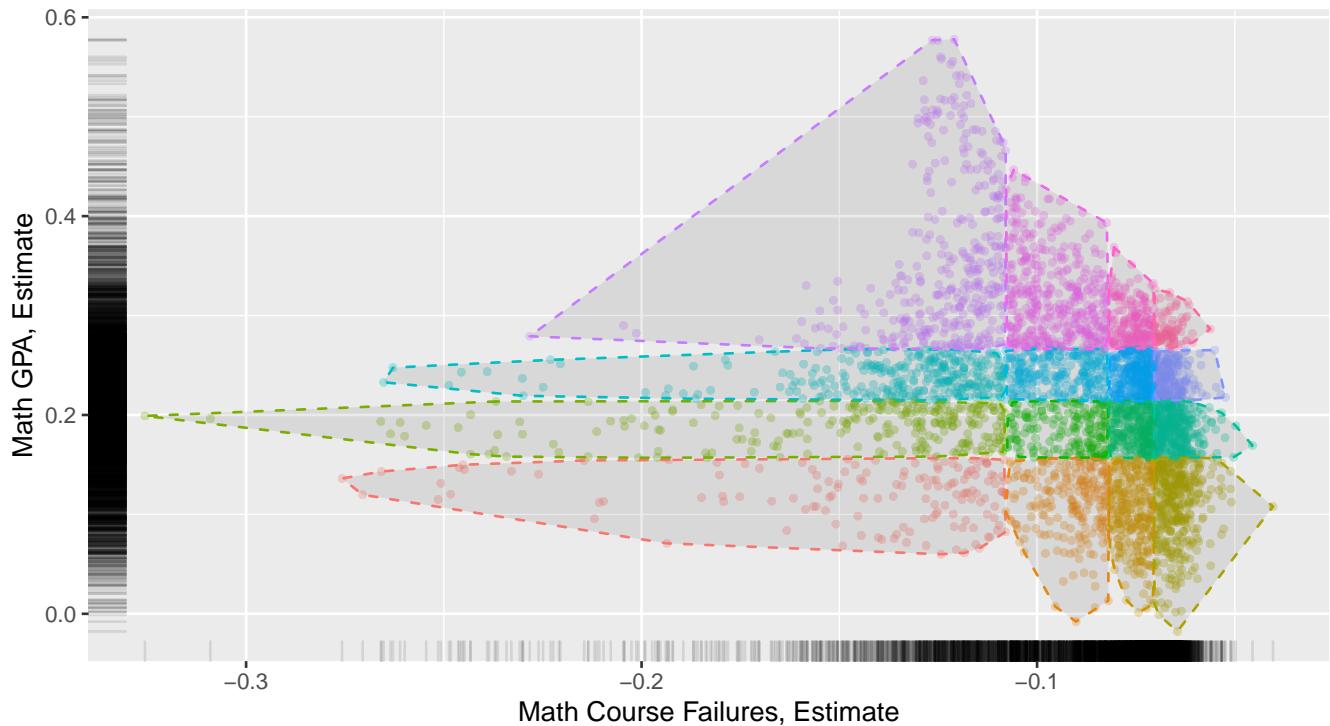
PTE Decile Plot: Math GPA vs. Math Course Failures



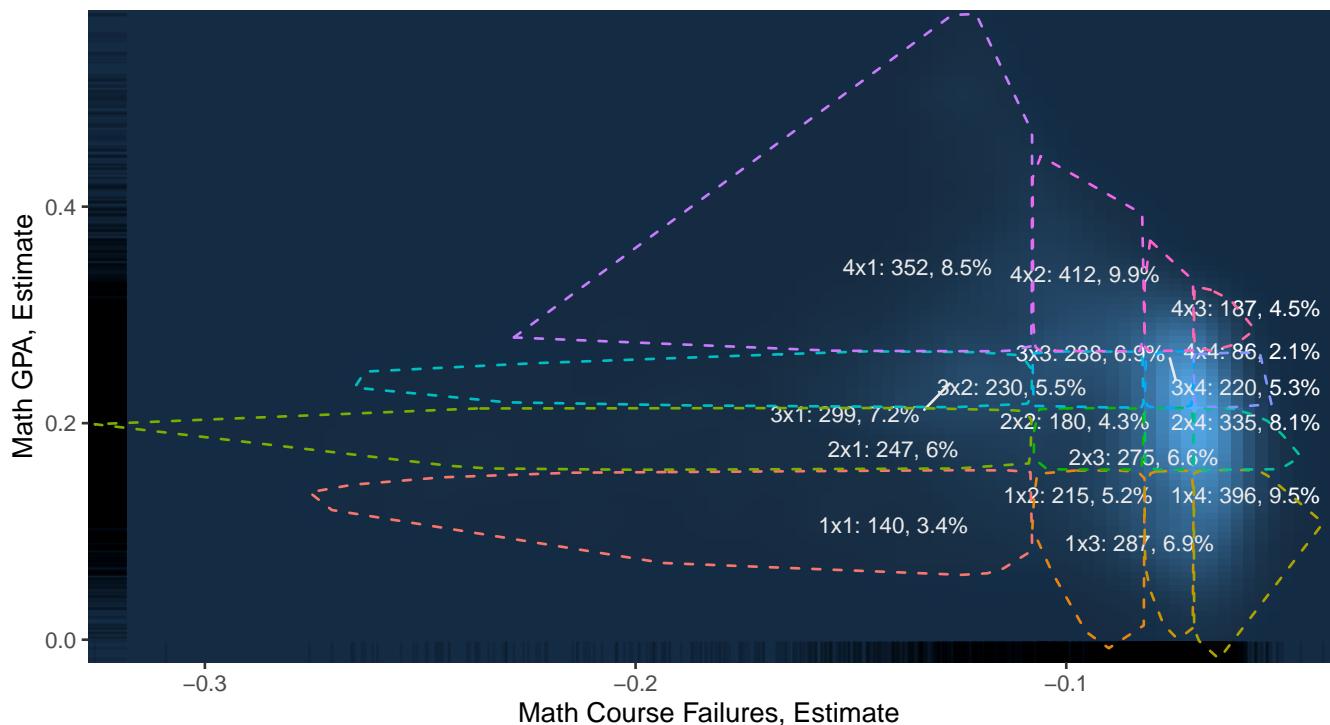
PTE Decile Plot: Math GPA vs. Math Course Failures



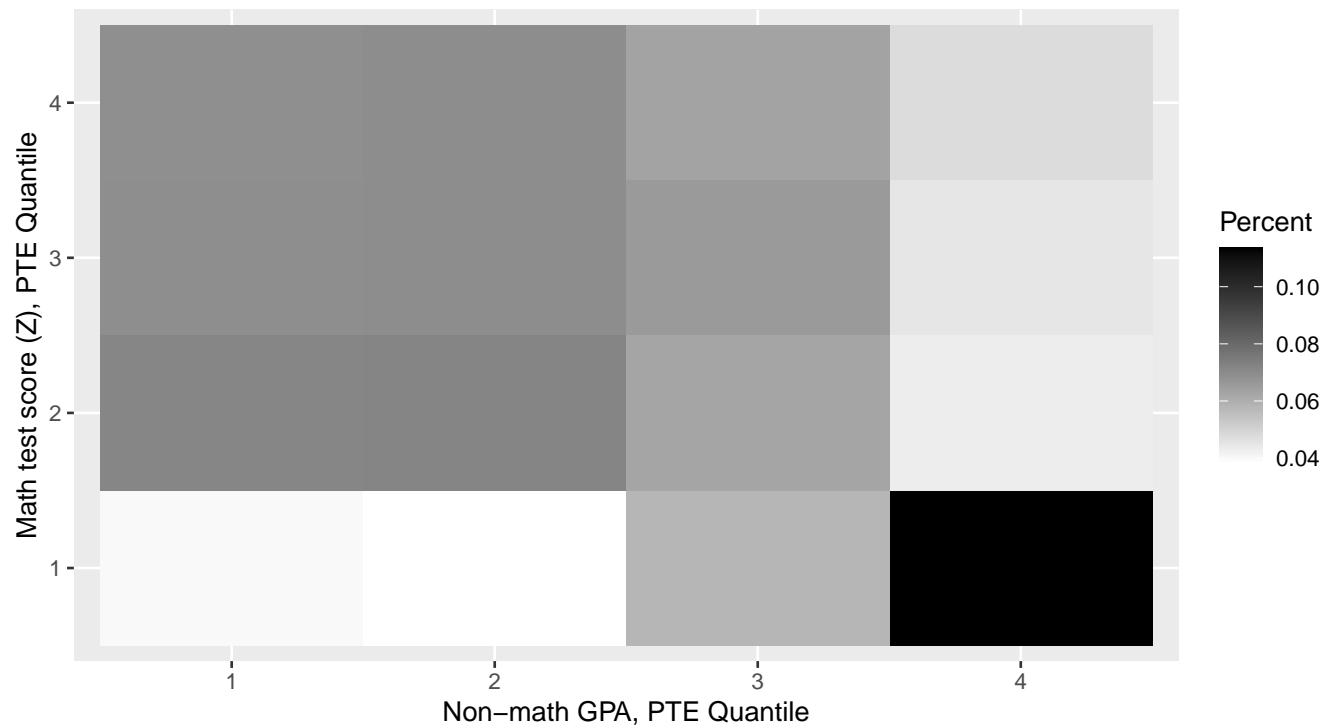
PTE Scatterplot: Math GPA vs Math Course Failures



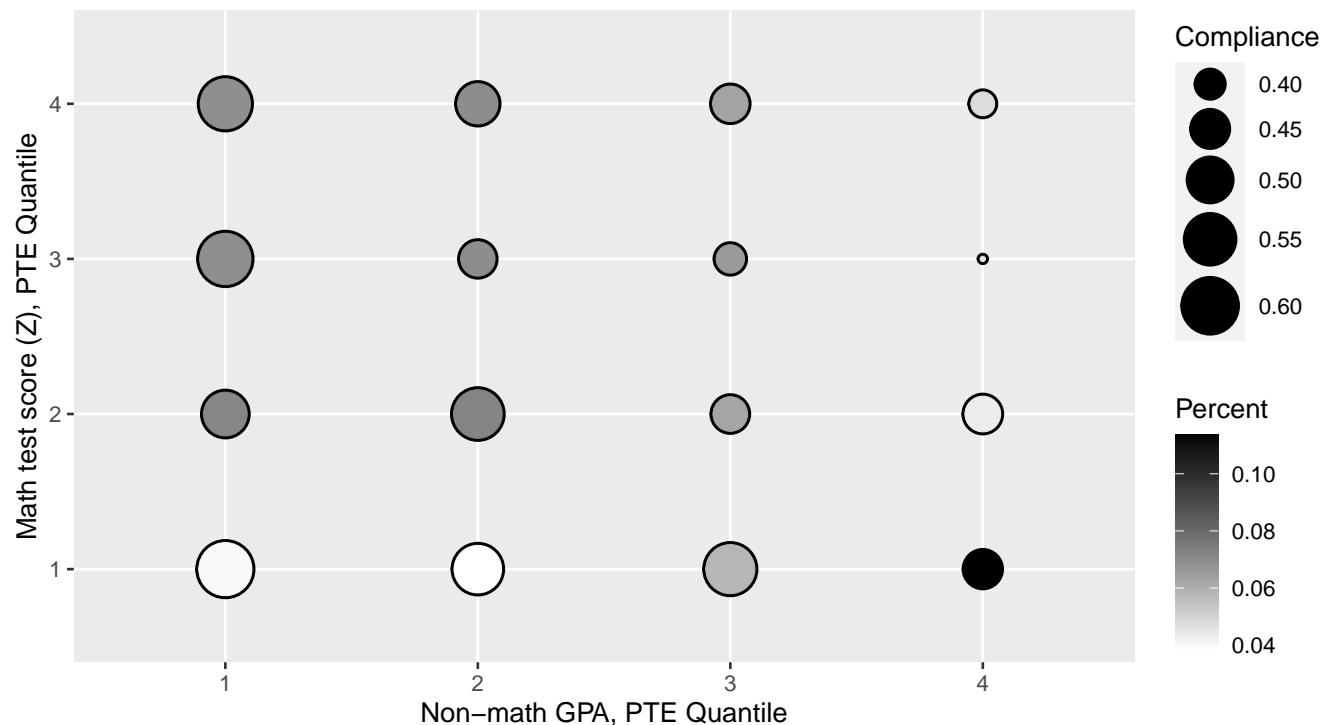
PTE Densities: Math GPA vs Math Course Failures



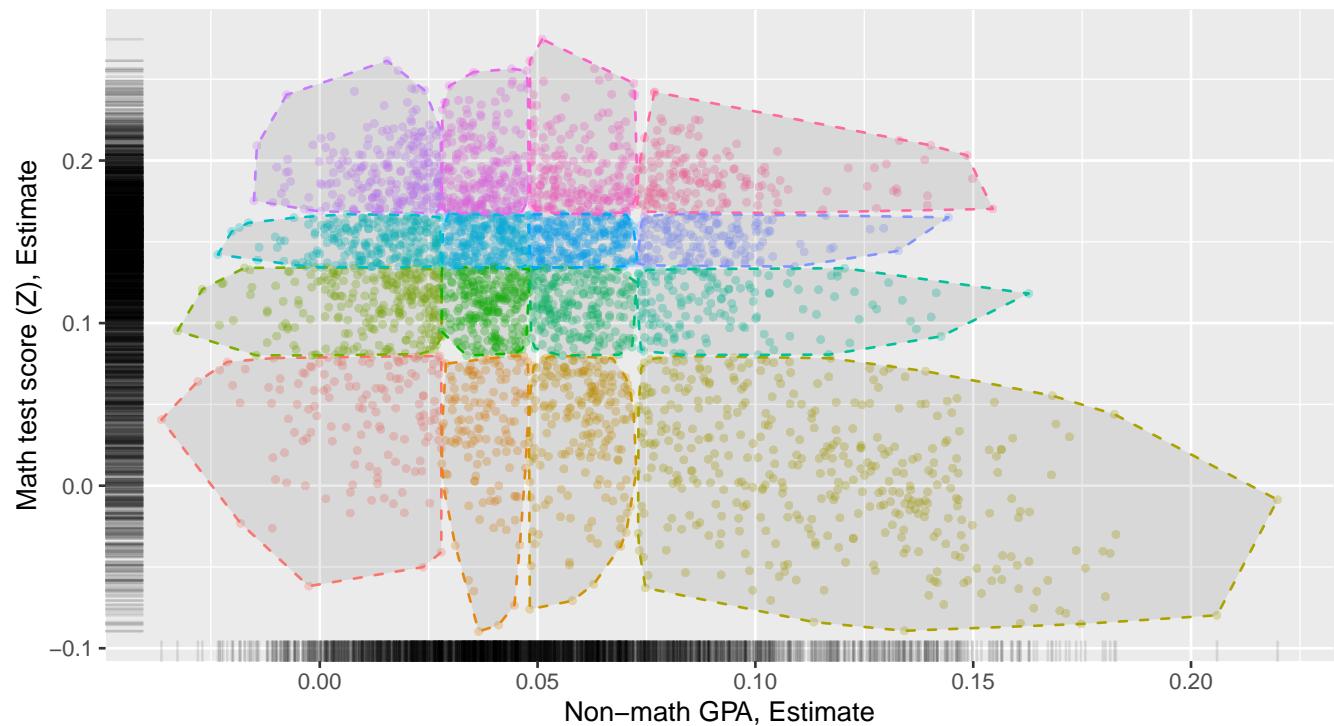
PTE Decile Plot: Math test score (Z) vs. Non–math GPA



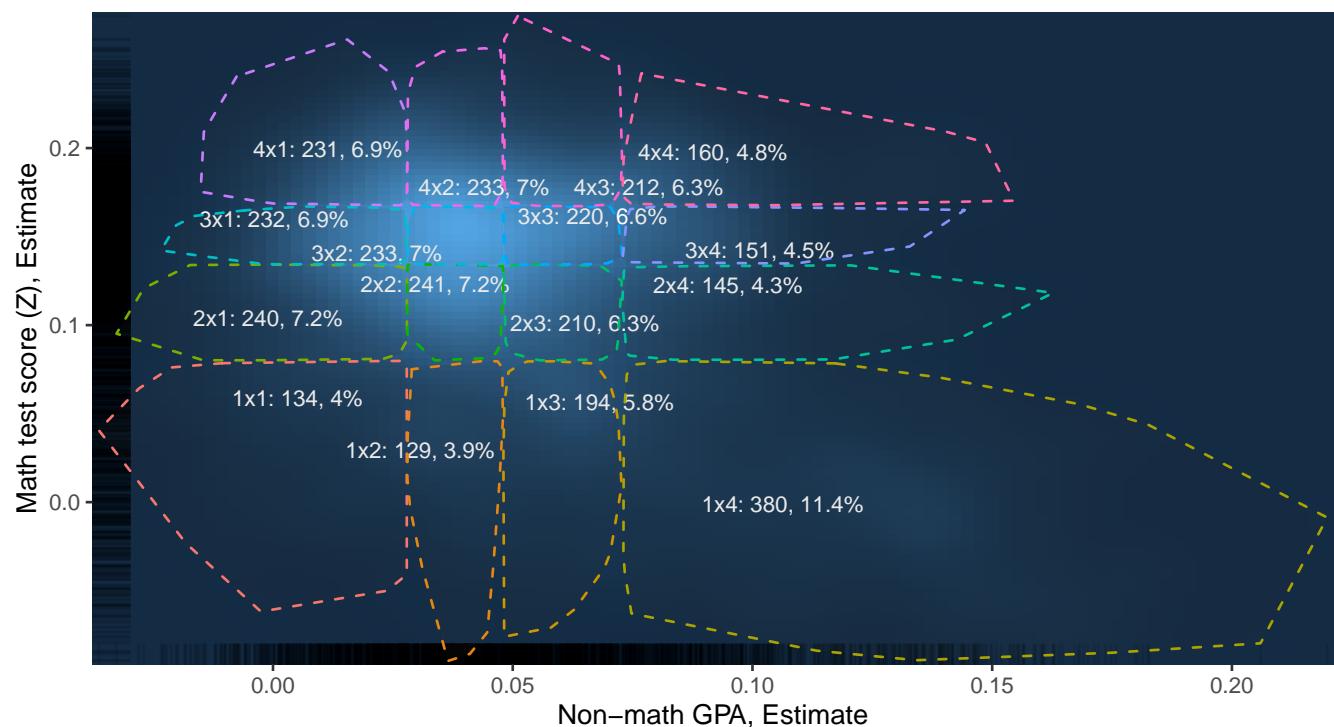
PTE Decile Plot: Math test score (Z) vs. Non–math GPA



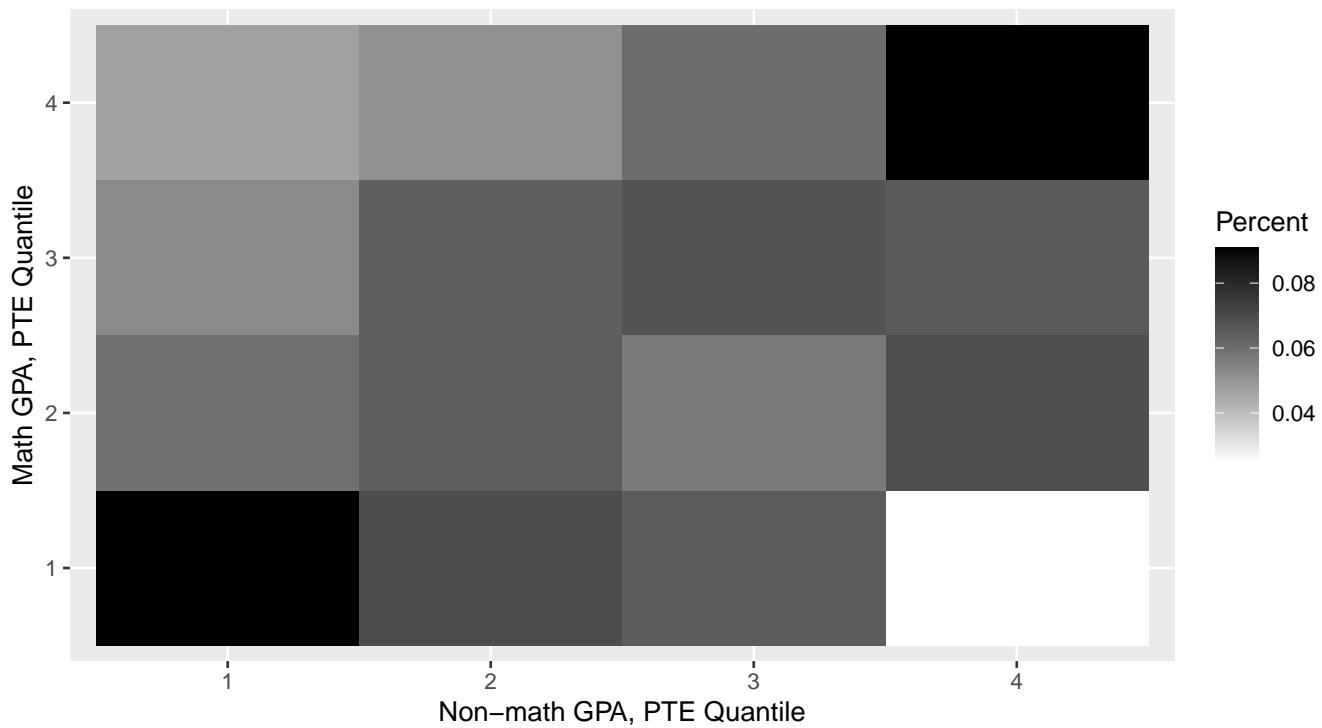
PTE Scatterplot: Math test score (Z) vs Non–math GPA



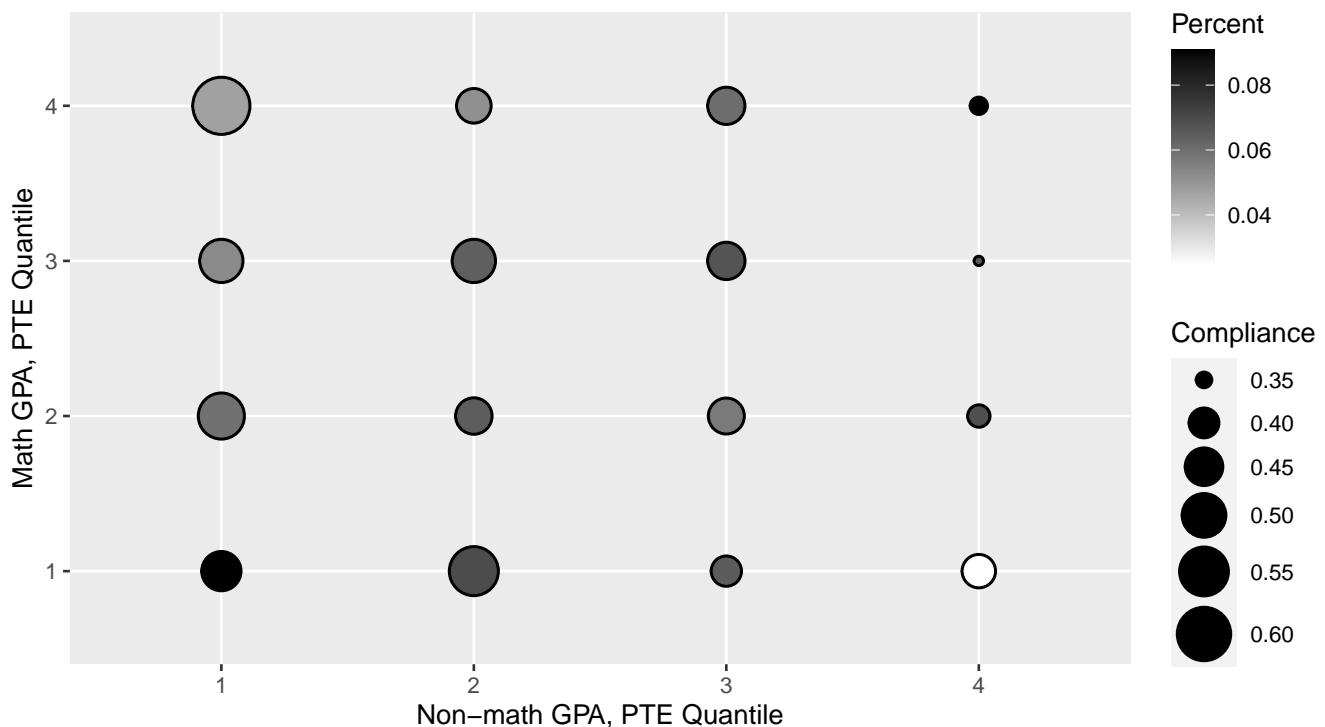
PTE Densities: Math test score (Z) vs Non–math GPA



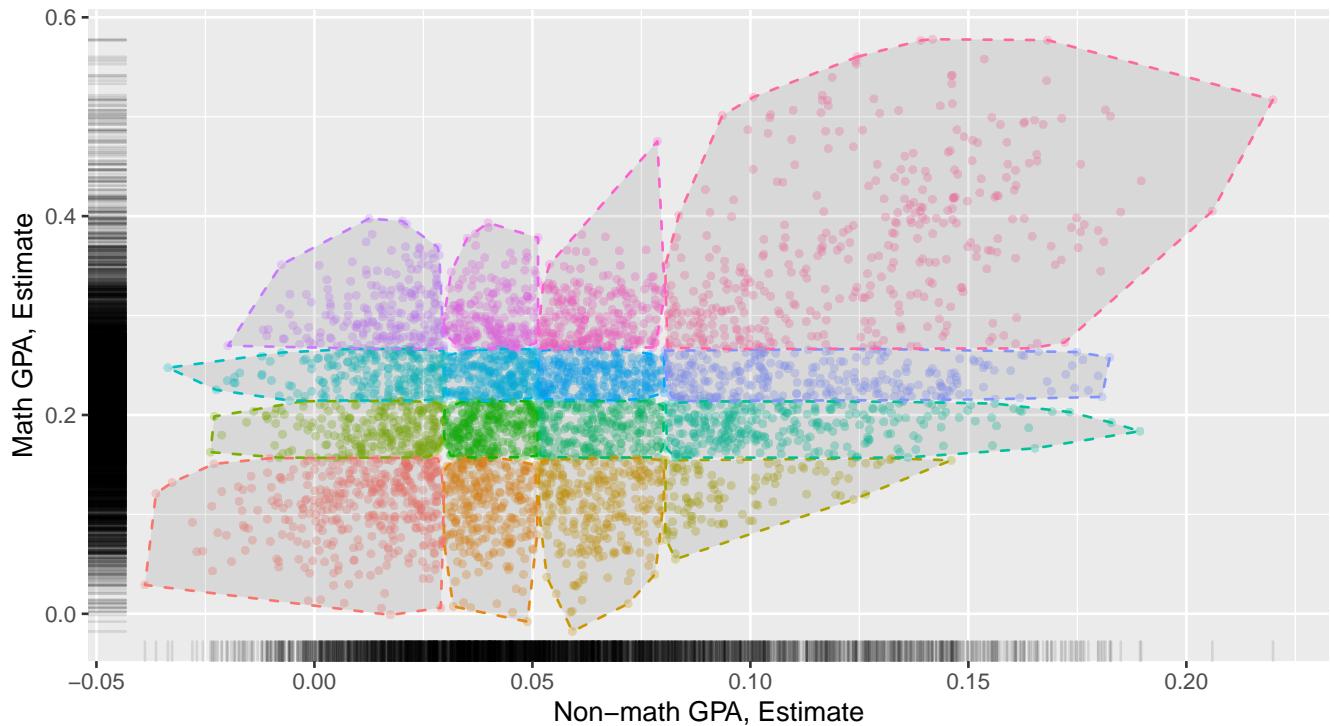
PTE Decile Plot: Math GPA vs. Non–math GPA



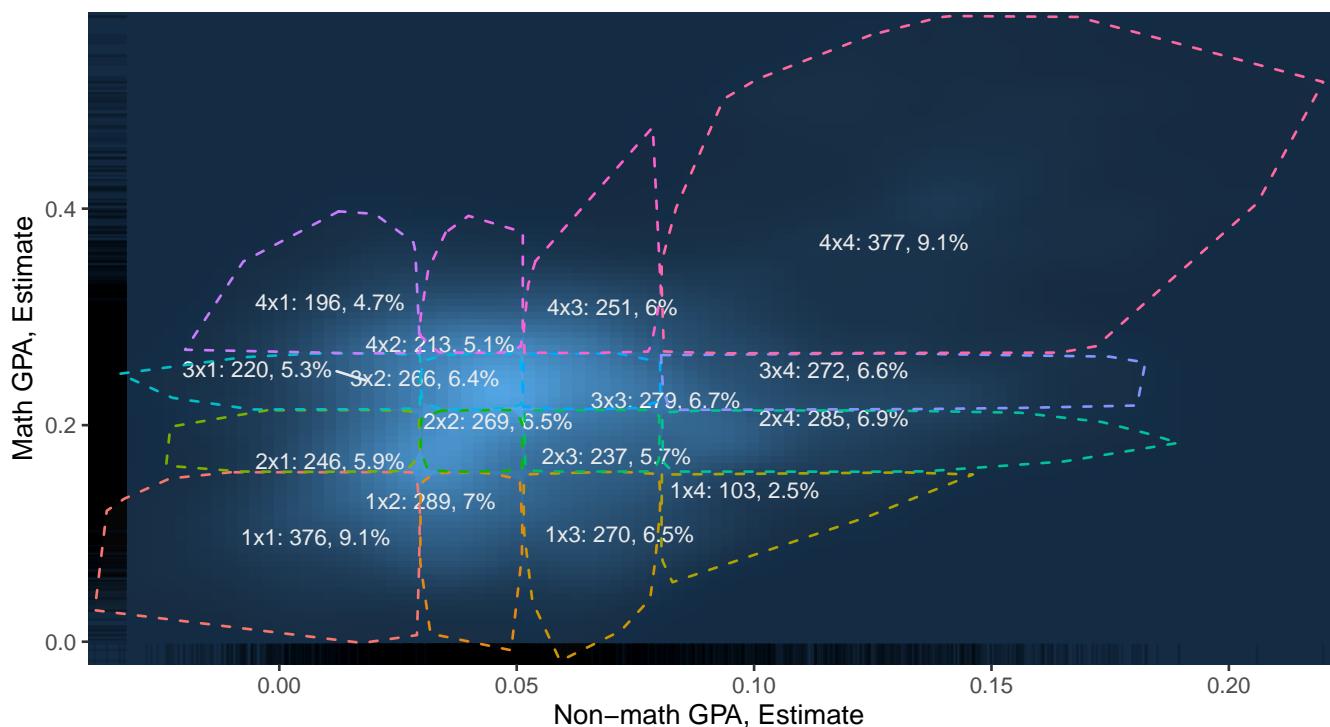
PTE Decile Plot: Math GPA vs. Non–math GPA



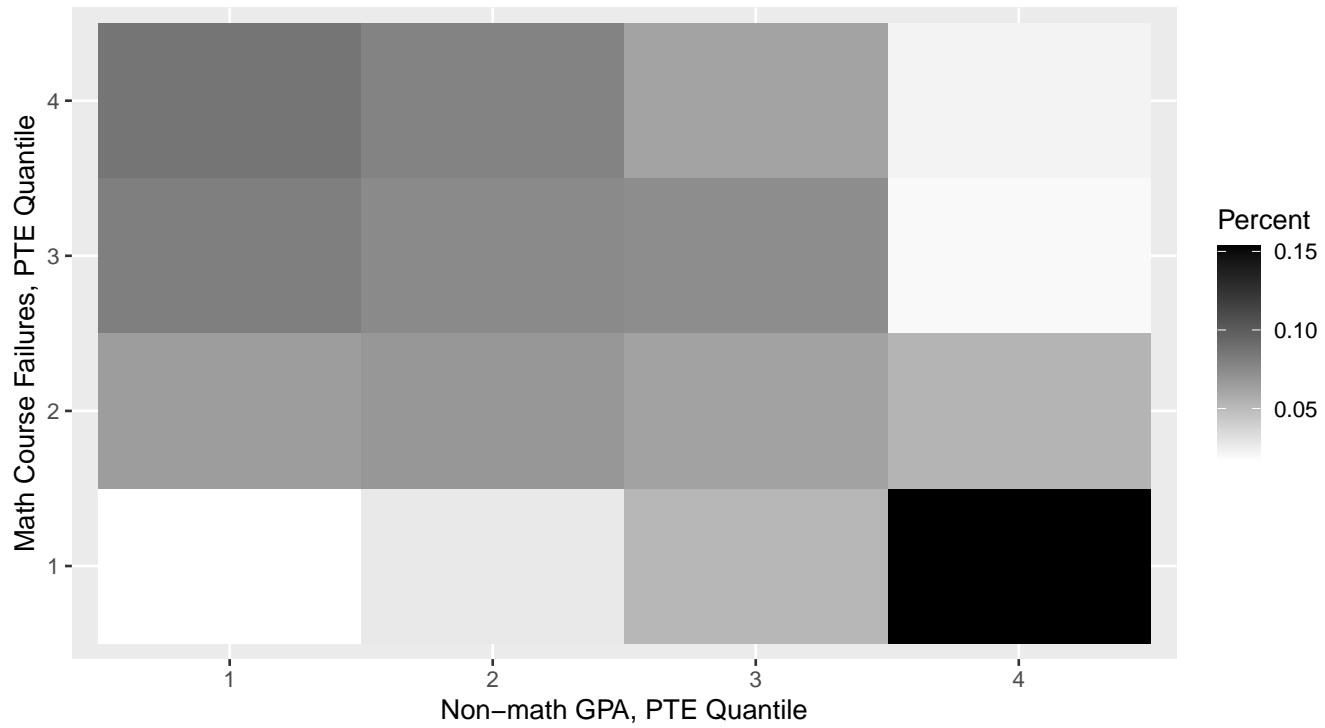
PTE Scatterplot: Math GPA vs Non–math GPA



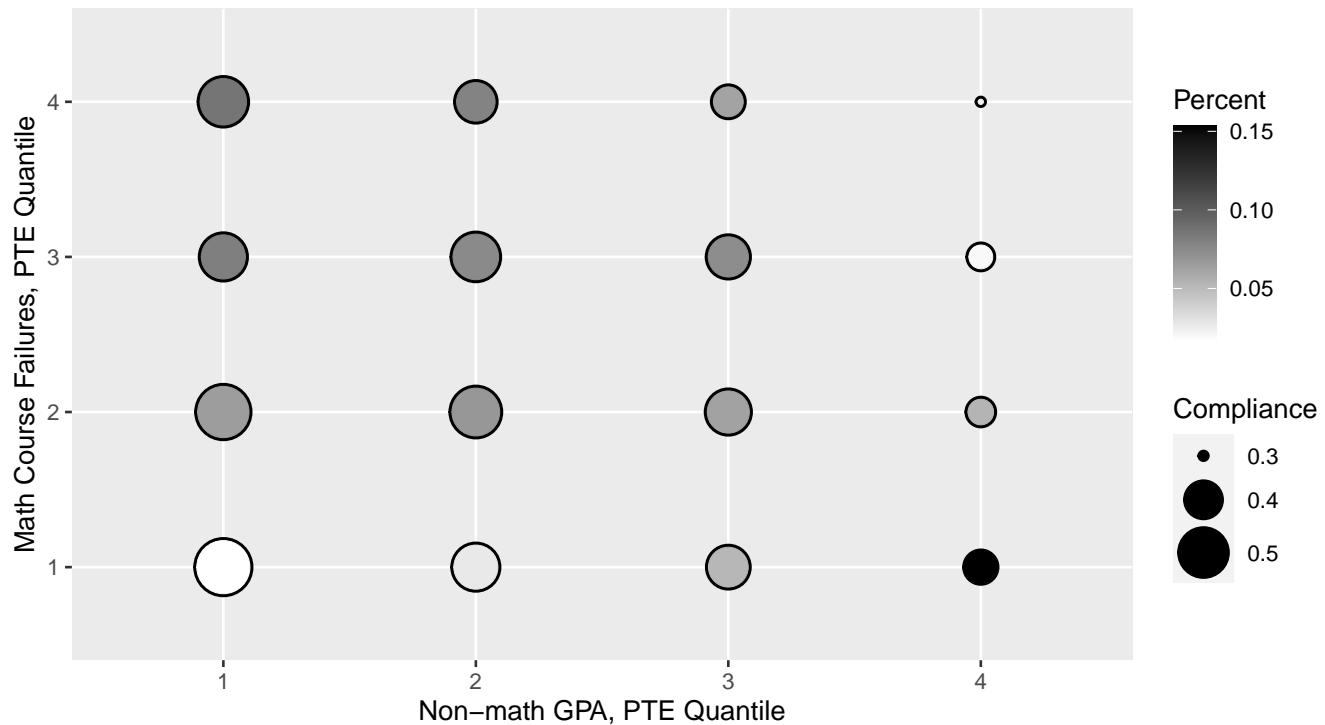
PTE Densities: Math GPA vs Non–math GPA



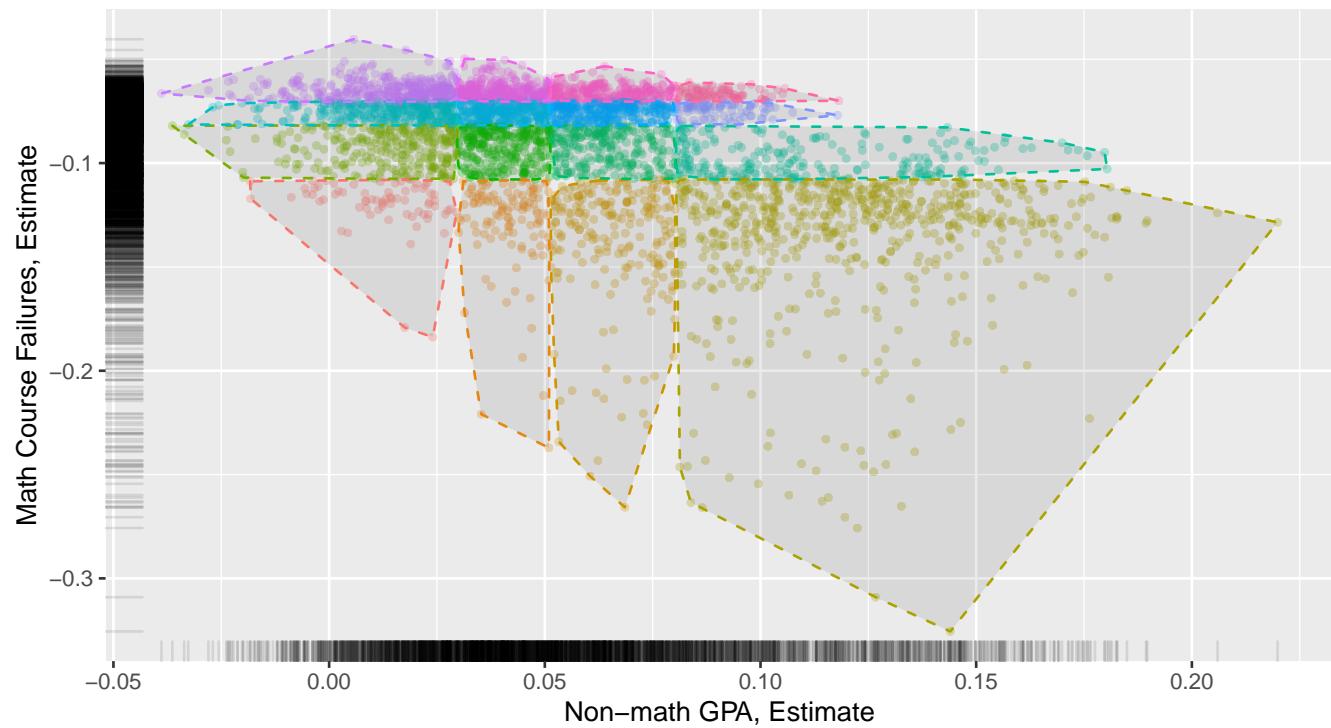
PTE Decile Plot: Math Course Failures vs. Non–math GPA



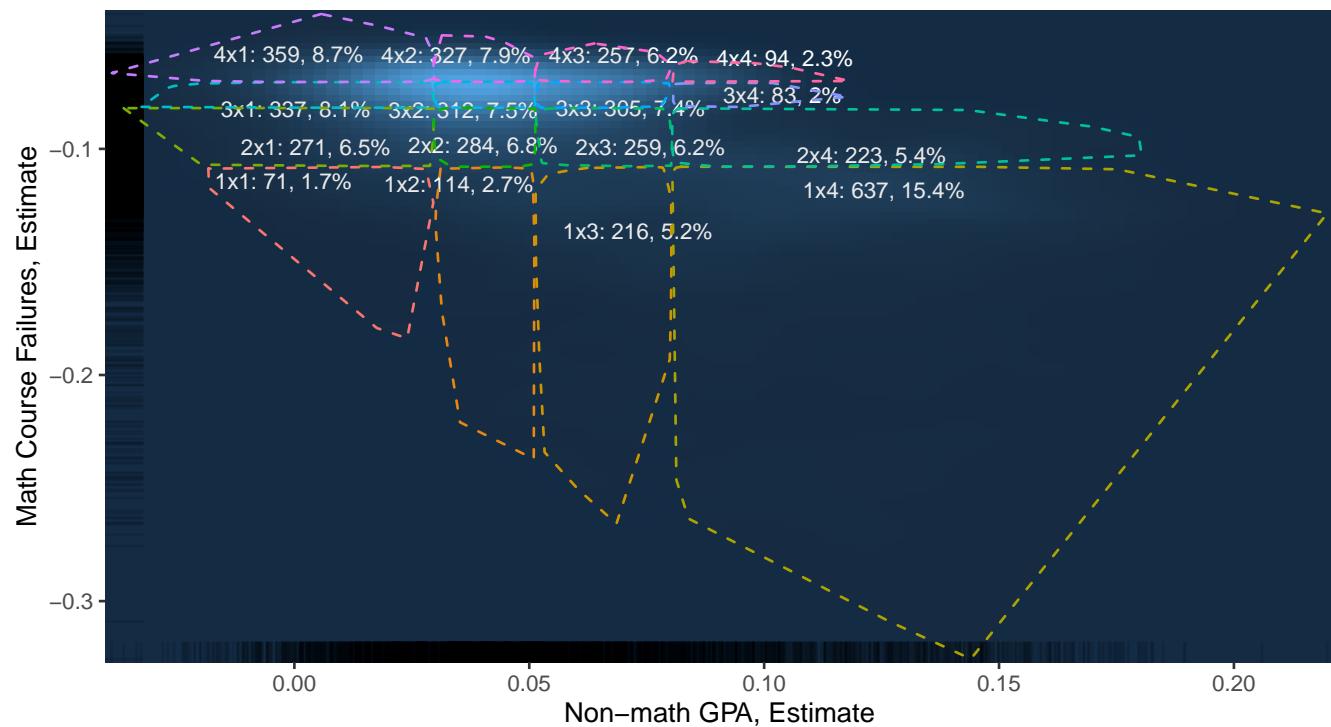
PTE Decile Plot: Math Course Failures vs. Non–math GPA



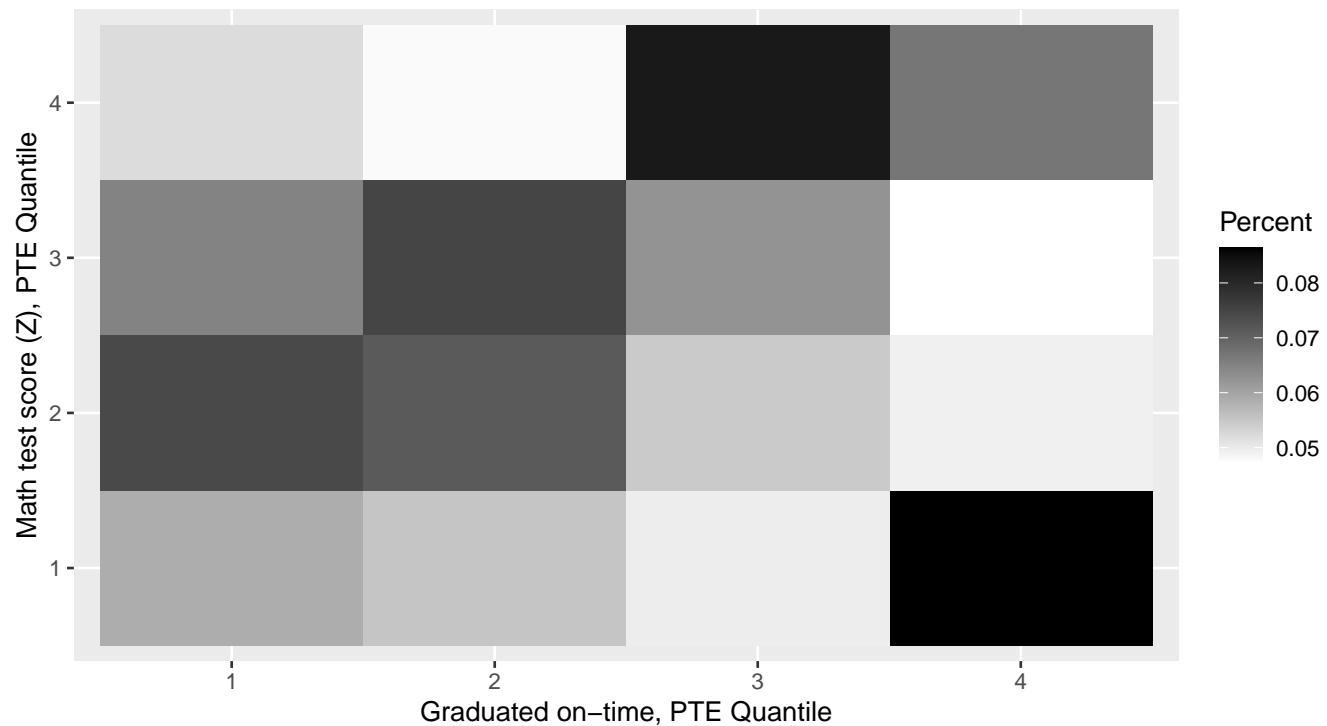
PTE Scatterplot: Math Course Failures vs Non–math GPA



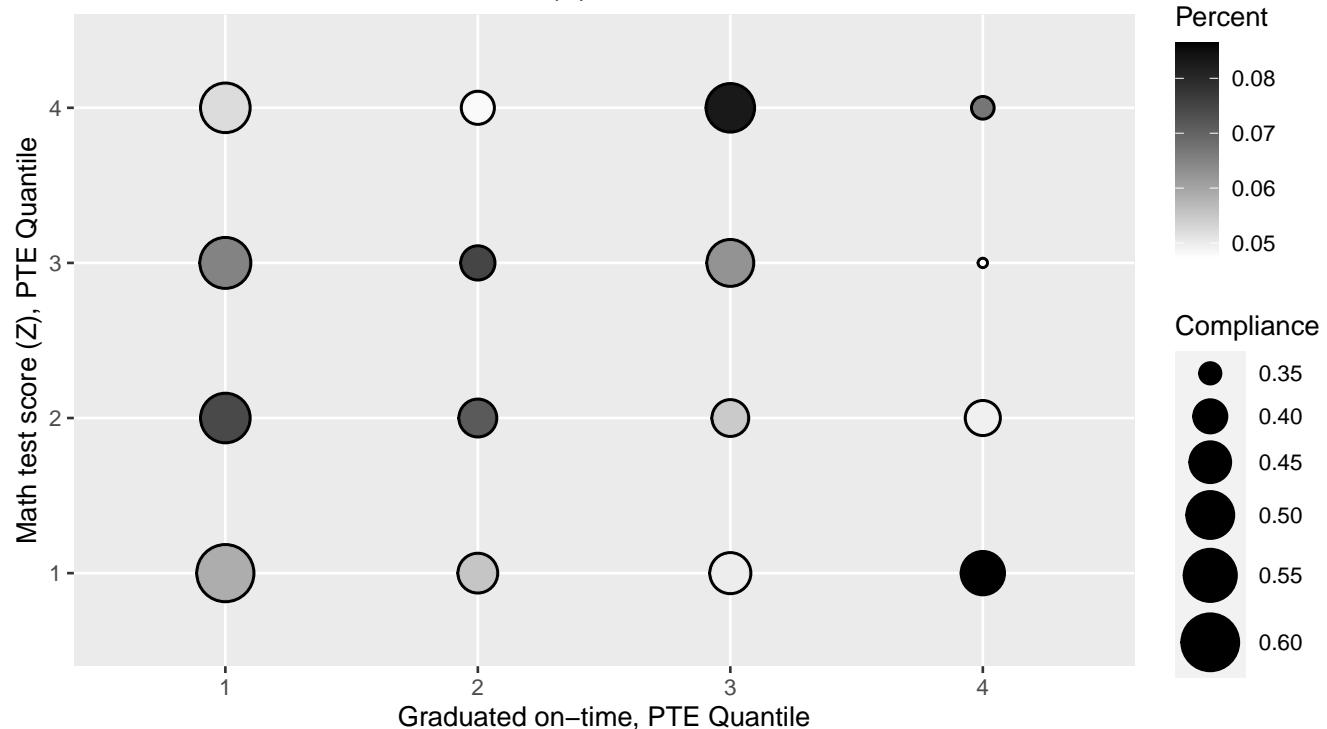
PTE Densities: Math Course Failures vs Non–math GPA



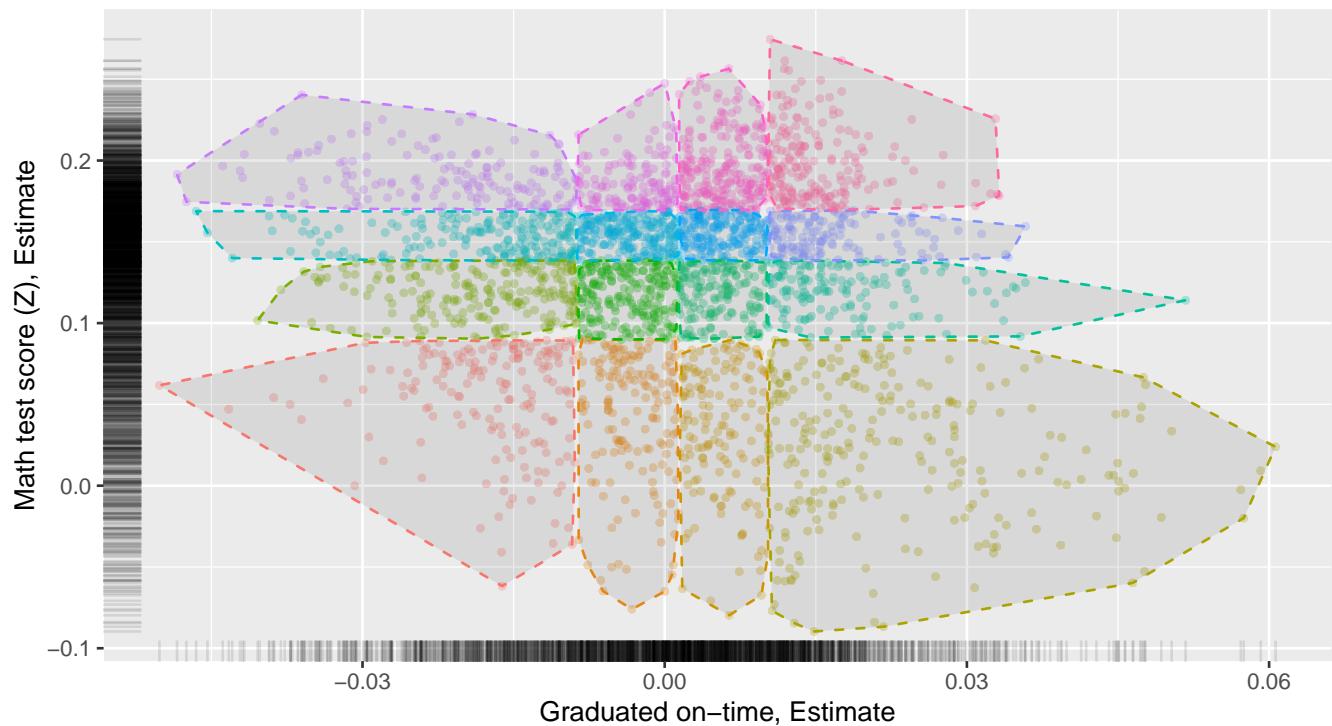
PTE Decile Plot: Math test score (Z) vs. Graduated on-time



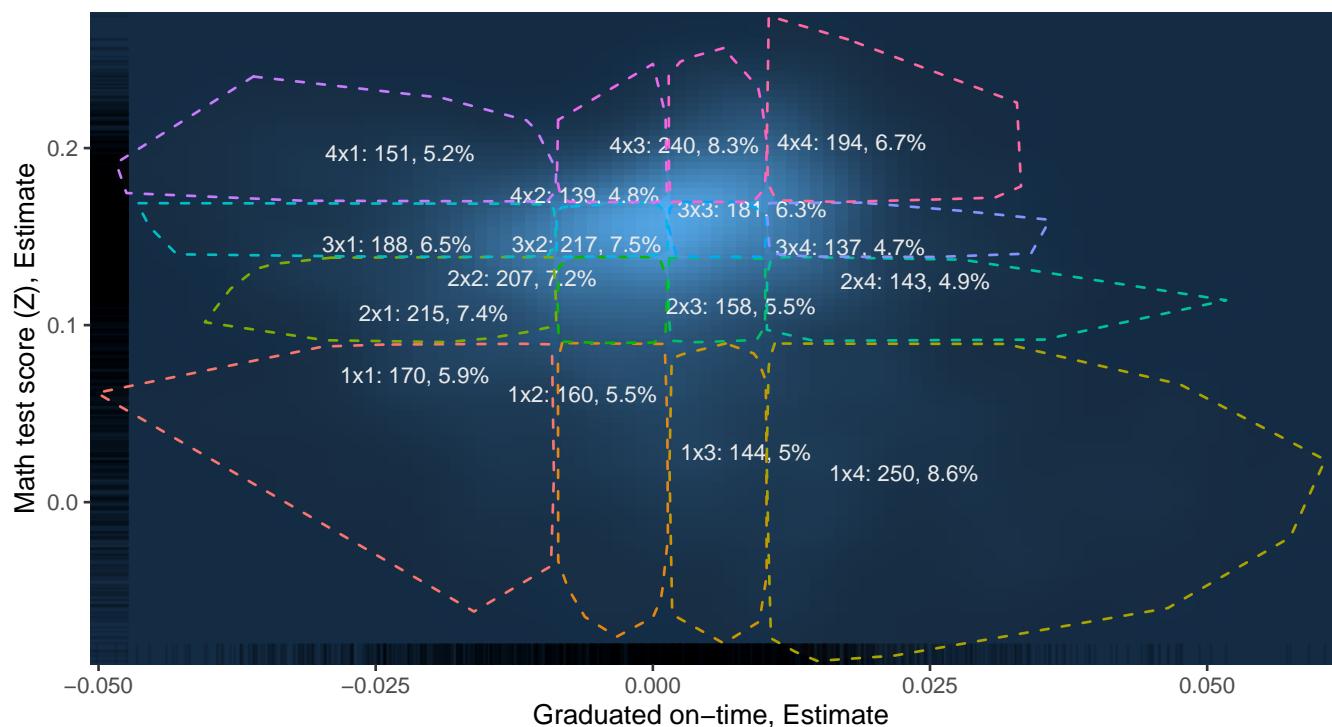
PTE Decile Plot: Math test score (Z) vs. Graduated on-time



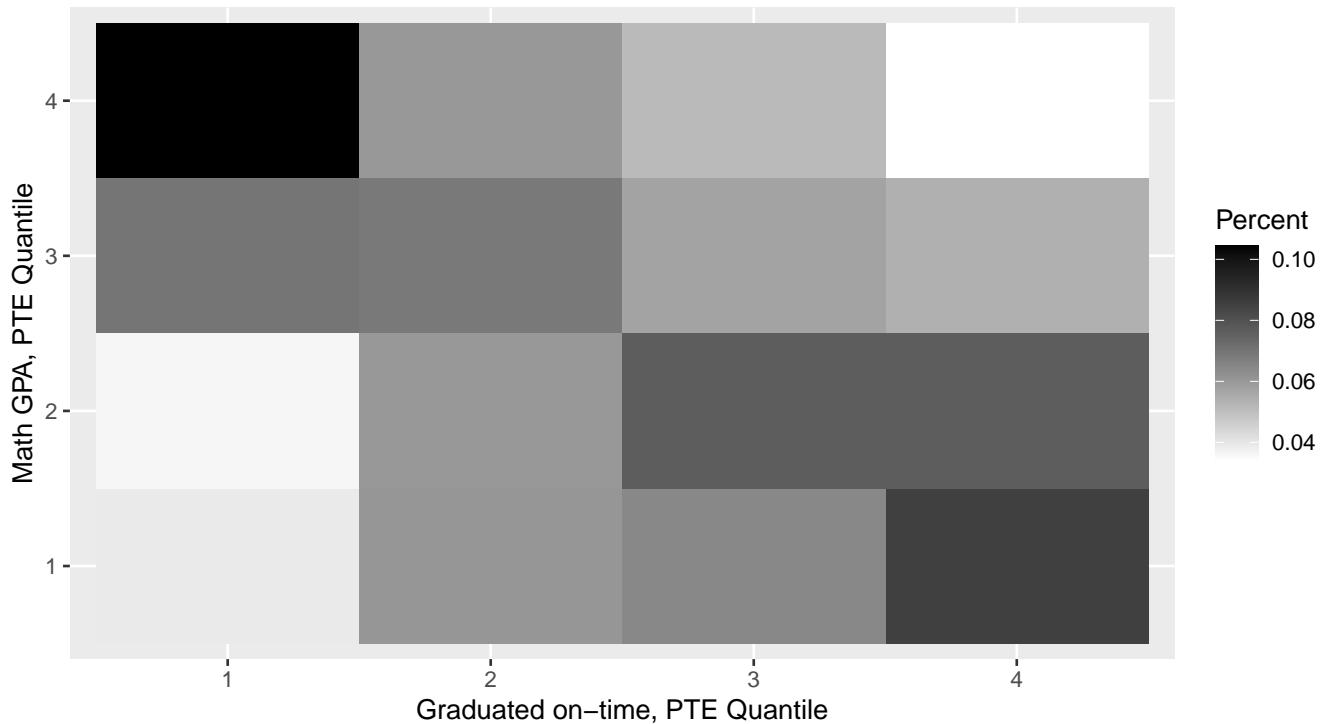
PTE Scatterplot: Math test score (Z) vs Graduated on-time



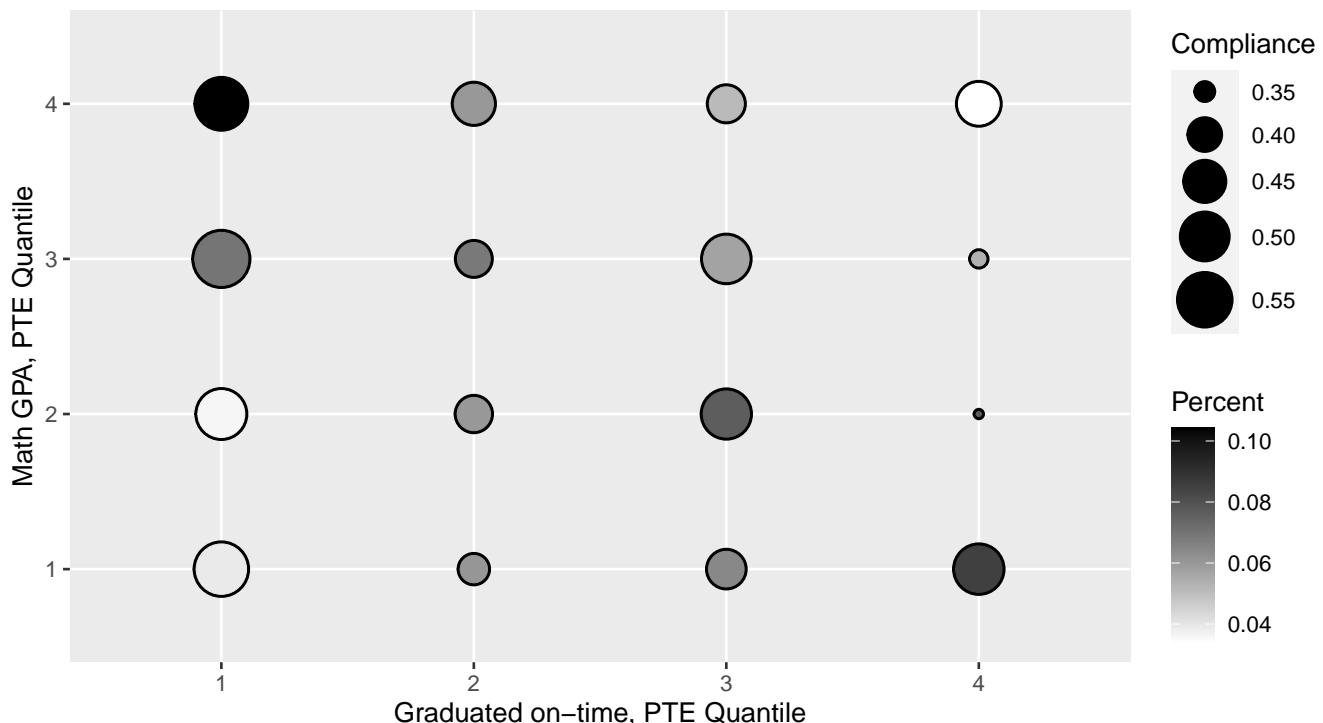
PTE Densities: Math test score (Z) vs Graduated on-time



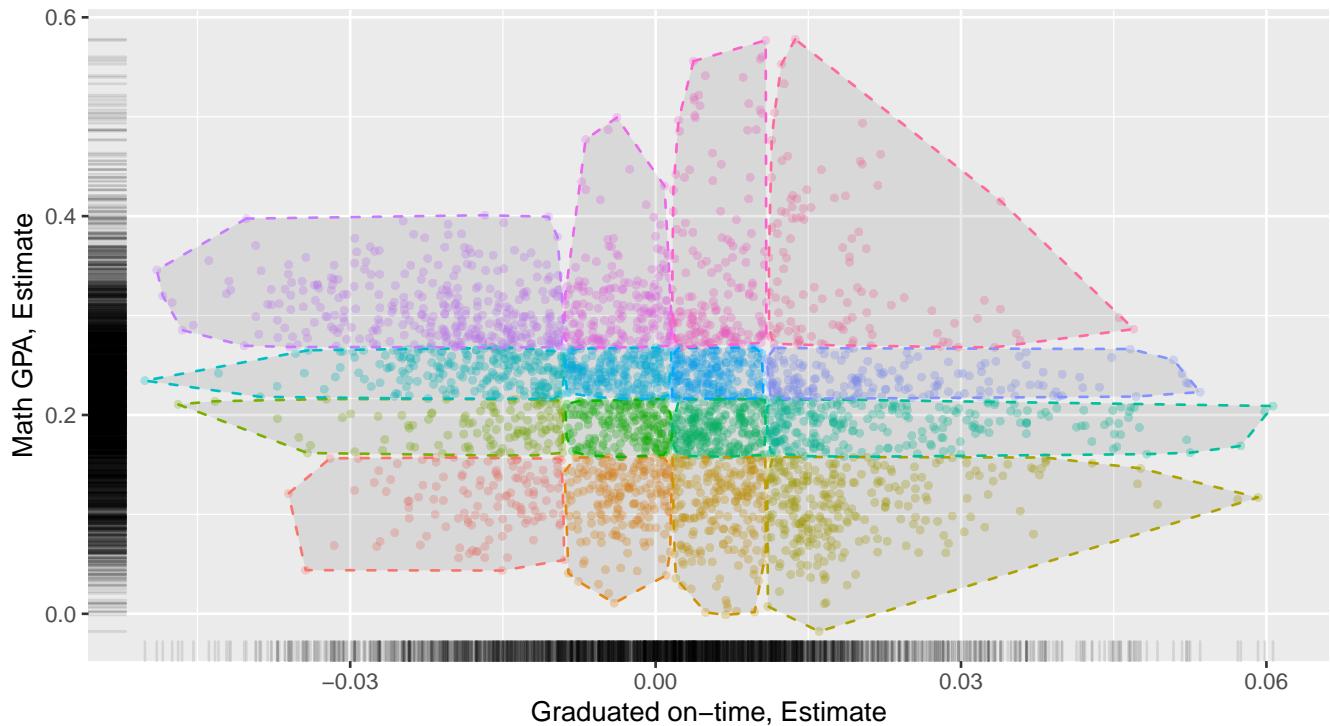
PTE Decile Plot: Math GPA vs. Graduated on-time



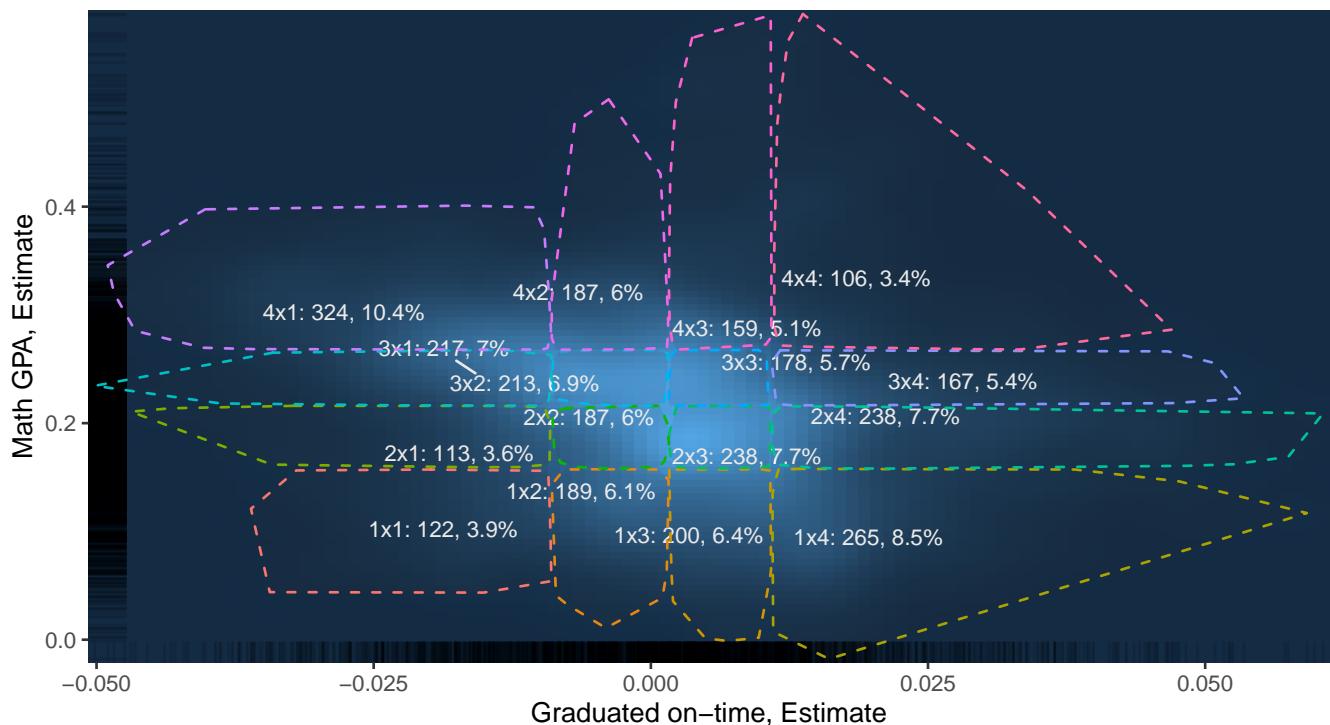
PTE Decile Plot: Math GPA vs. Graduated on-time



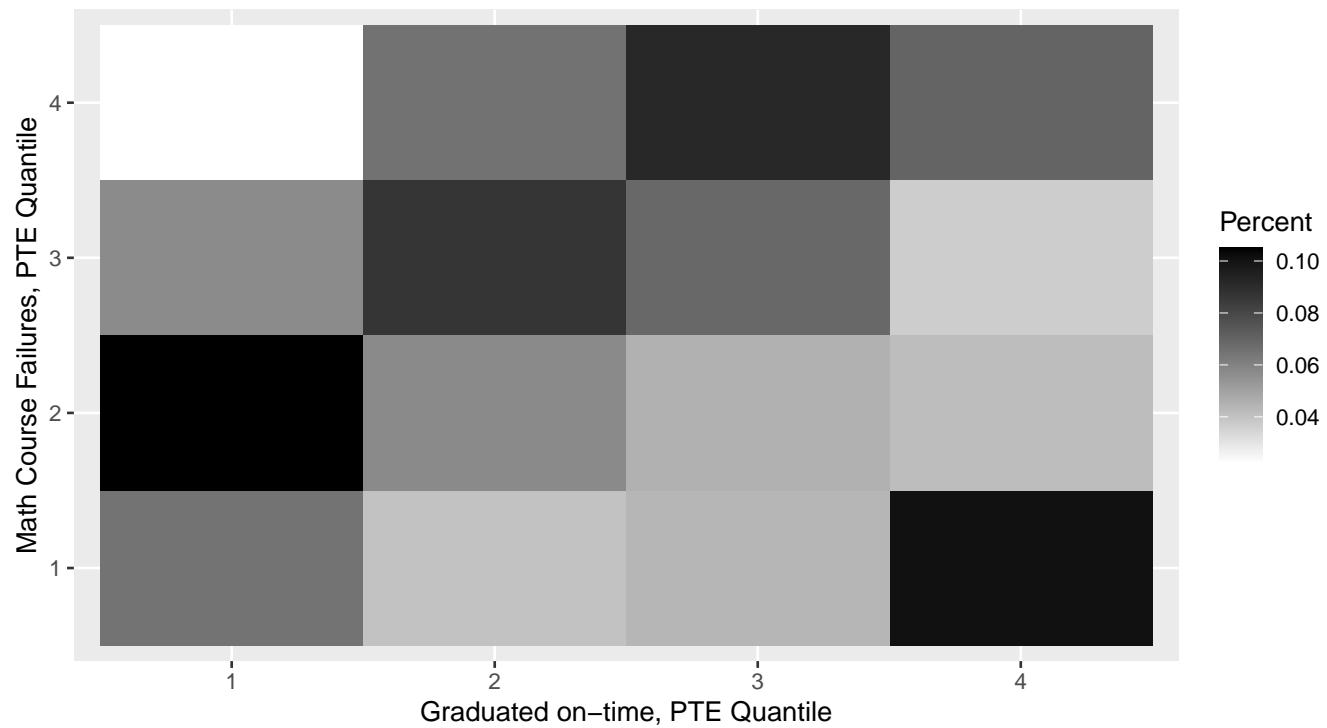
PTE Scatterplot: Math GPA vs Graduated on-time



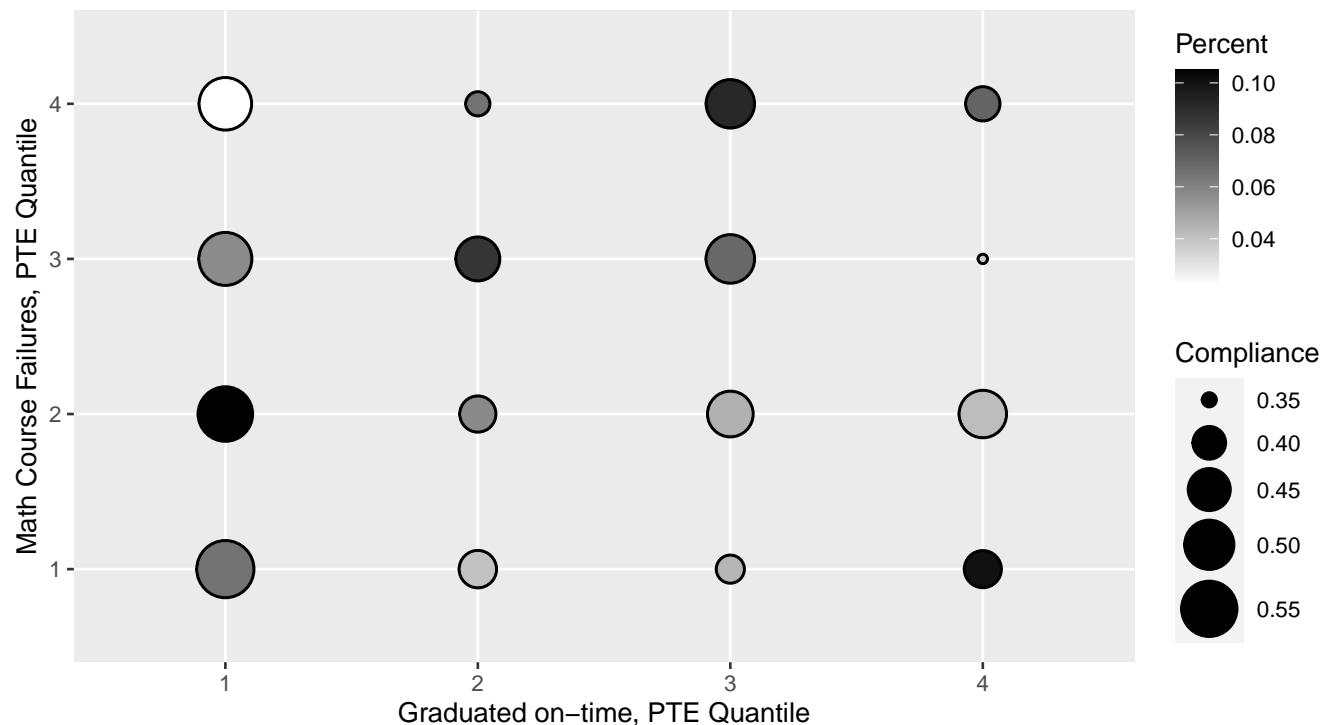
PTE Densities: Math GPA vs Graduated on-time



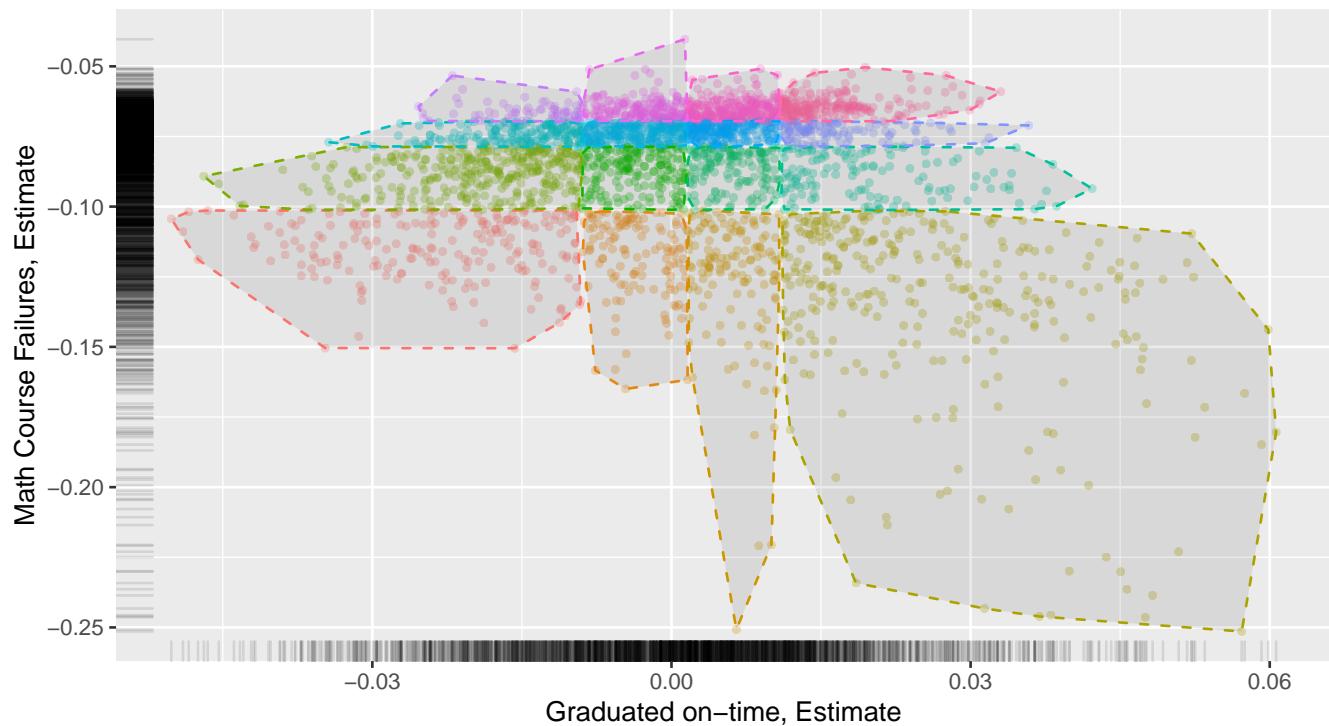
PTE Decile Plot: Math Course Failures vs. Graduated on-time



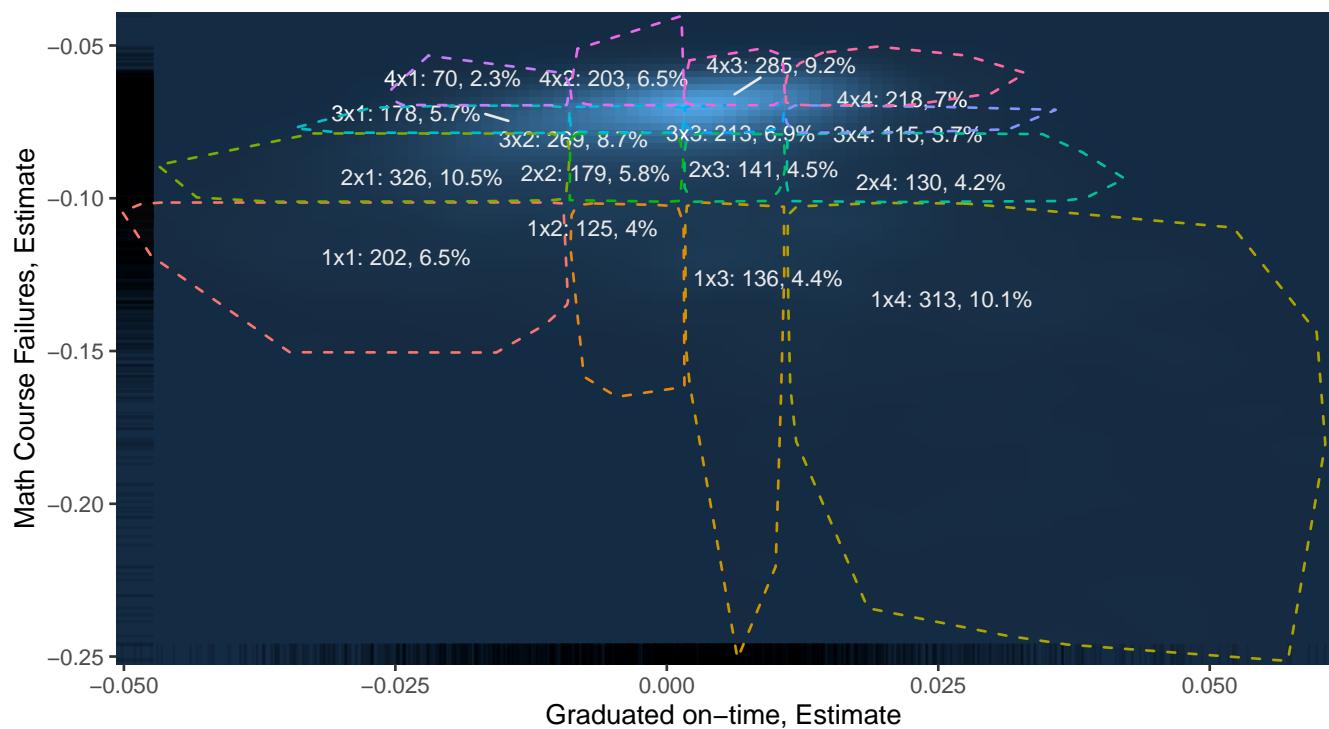
PTE Decile Plot: Math Course Failures vs. Graduated on-time



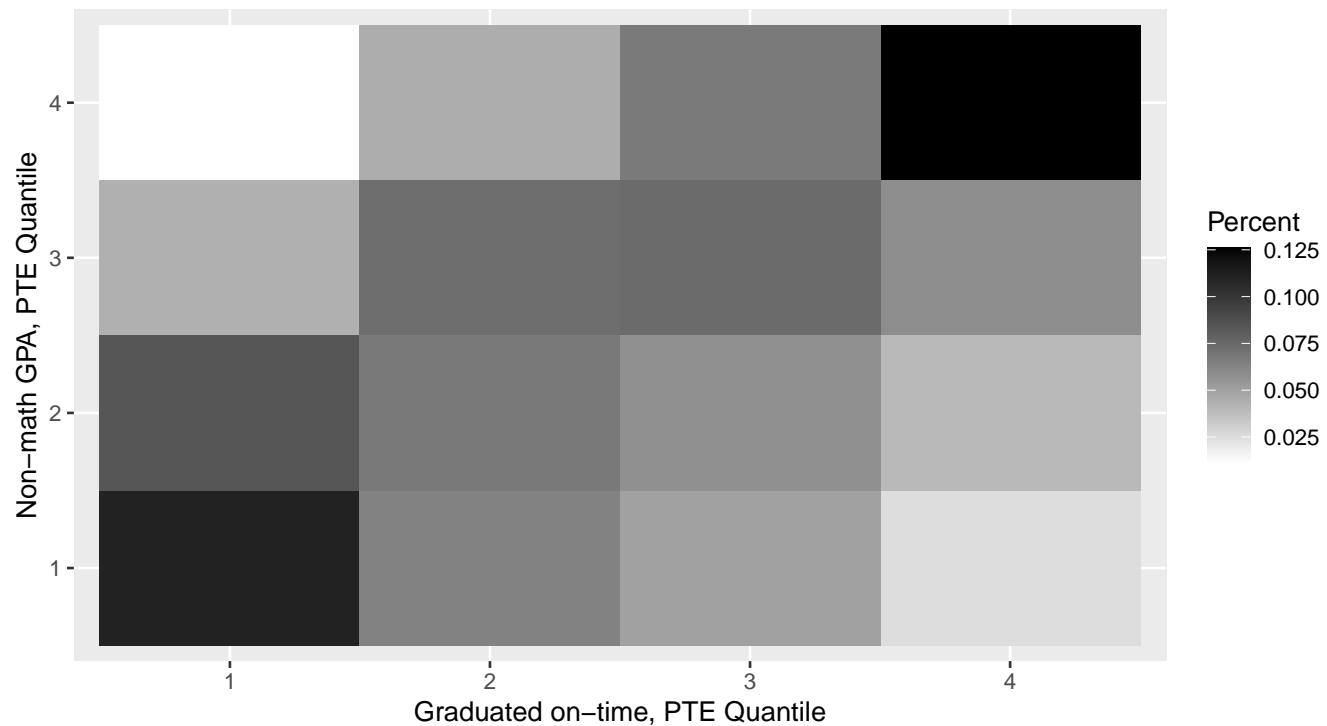
PTE Scatterplot: Math Course Failures vs Graduated on-time



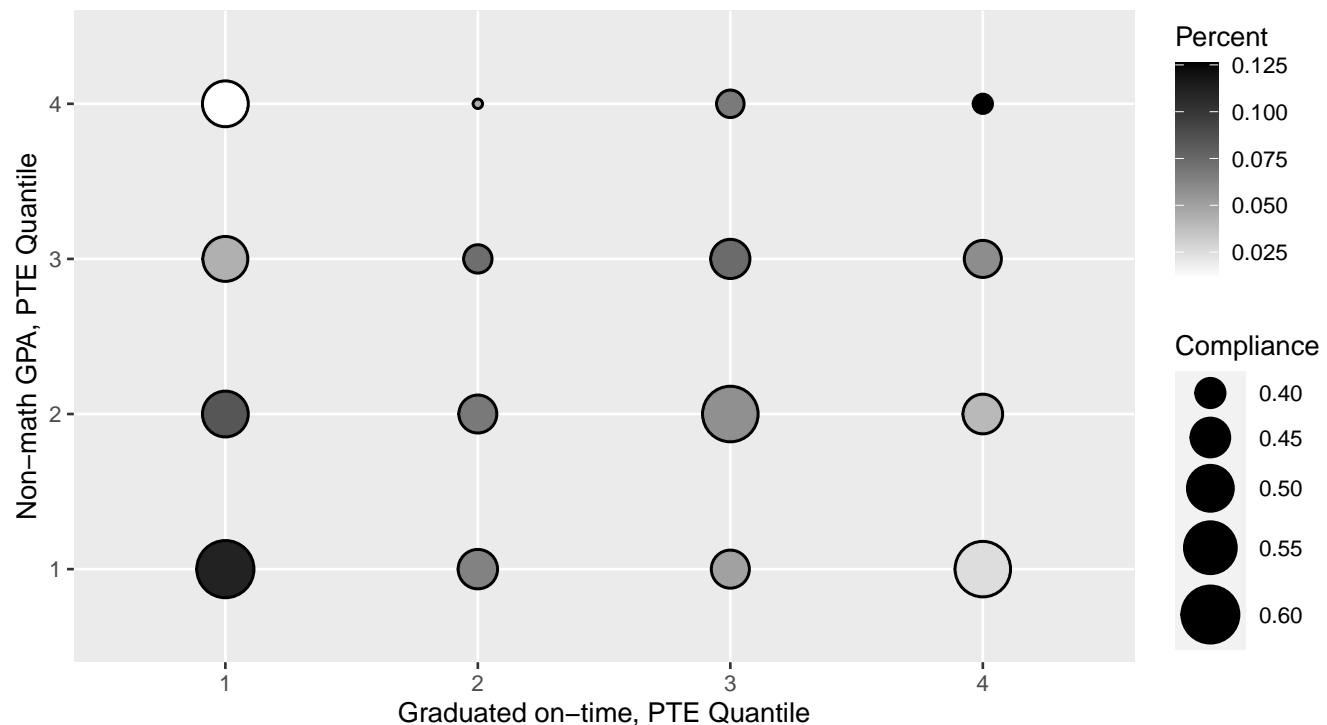
PTE Densities: Math Course Failures vs Graduated on-time



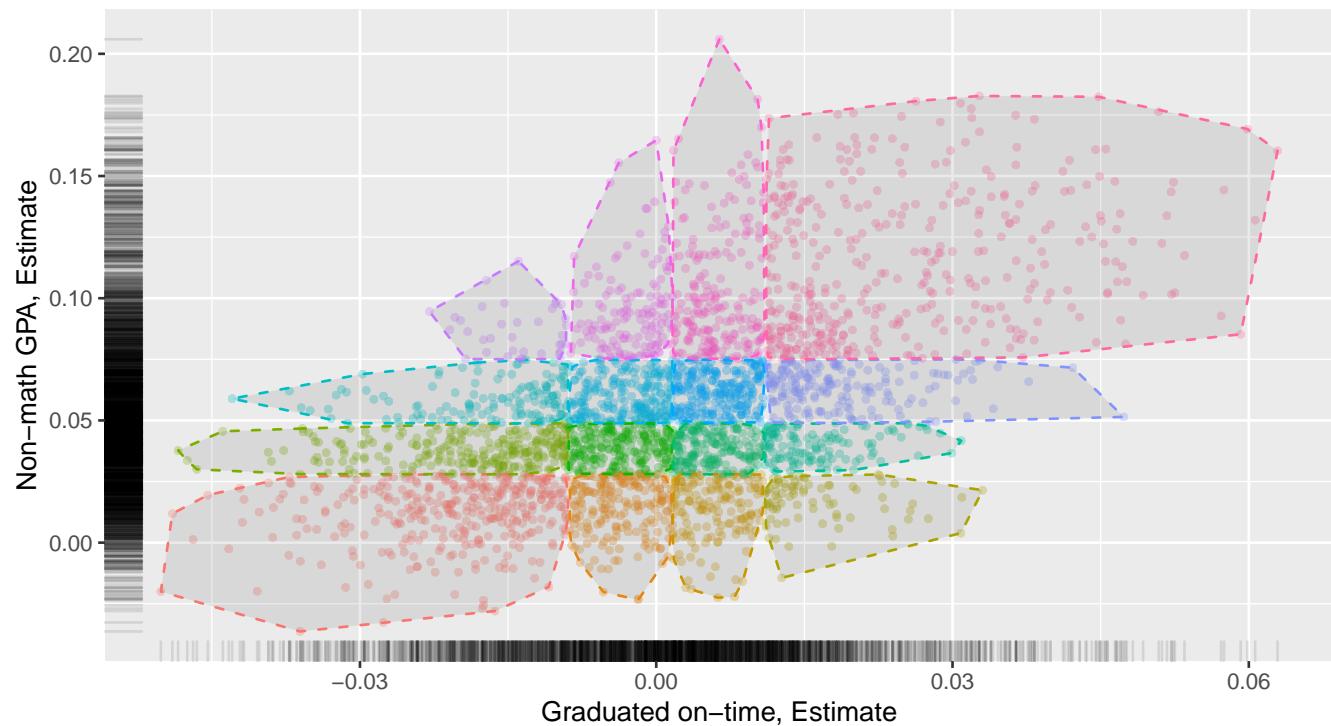
PTE Decile Plot: Non–math GPA vs. Graduated on–time



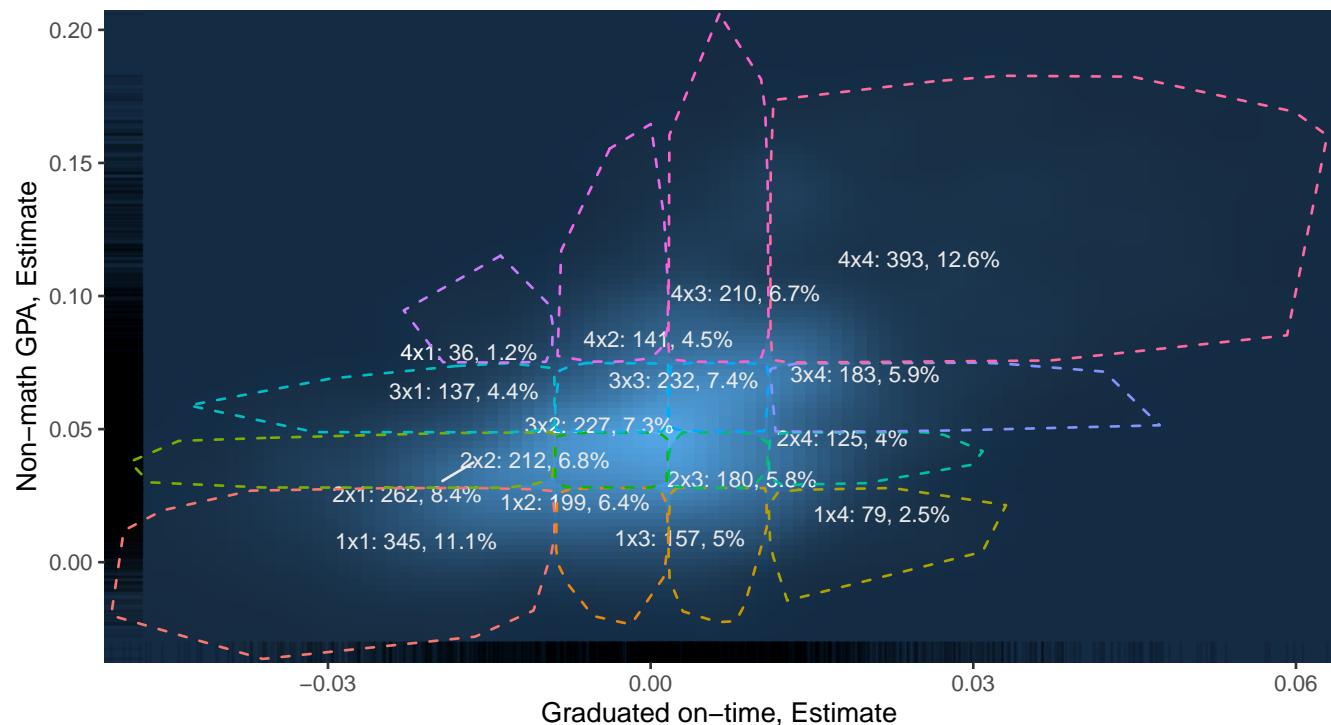
PTE Decile Plot: Non–math GPA vs. Graduated on–time



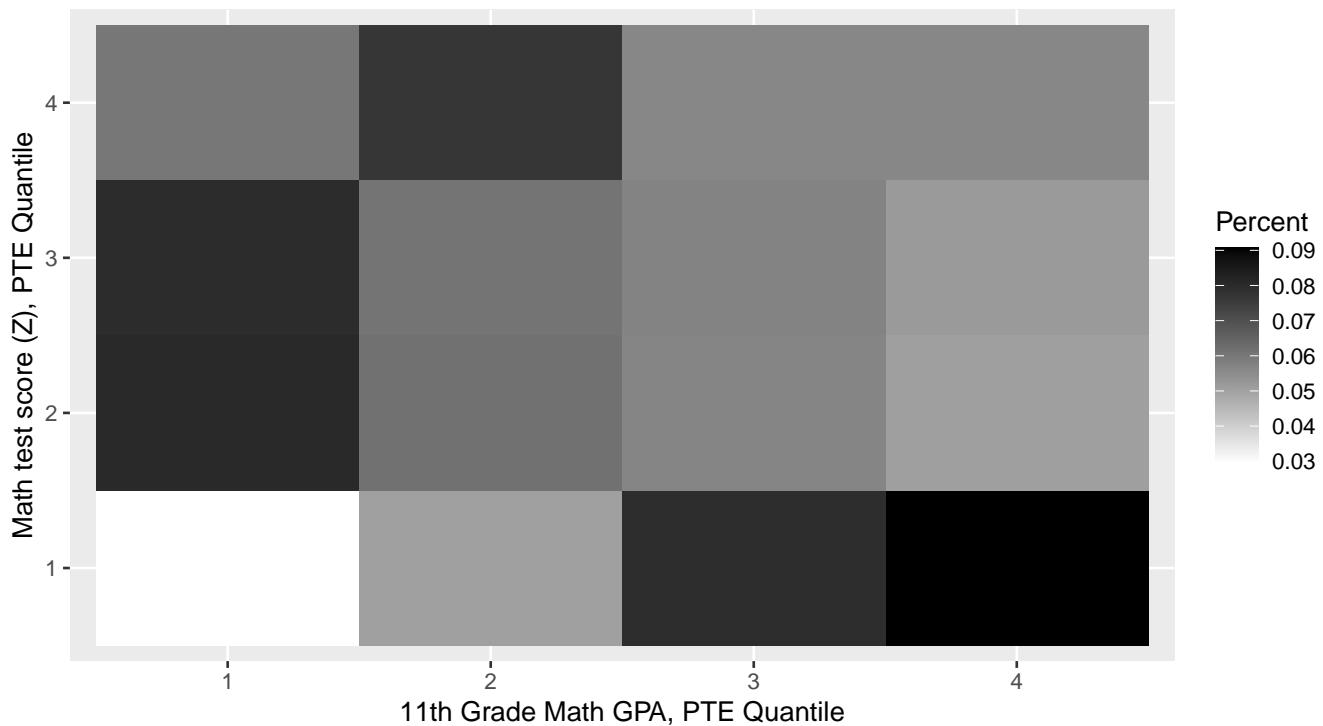
PTE Scatterplot: Non–math GPA vs Graduated on–time



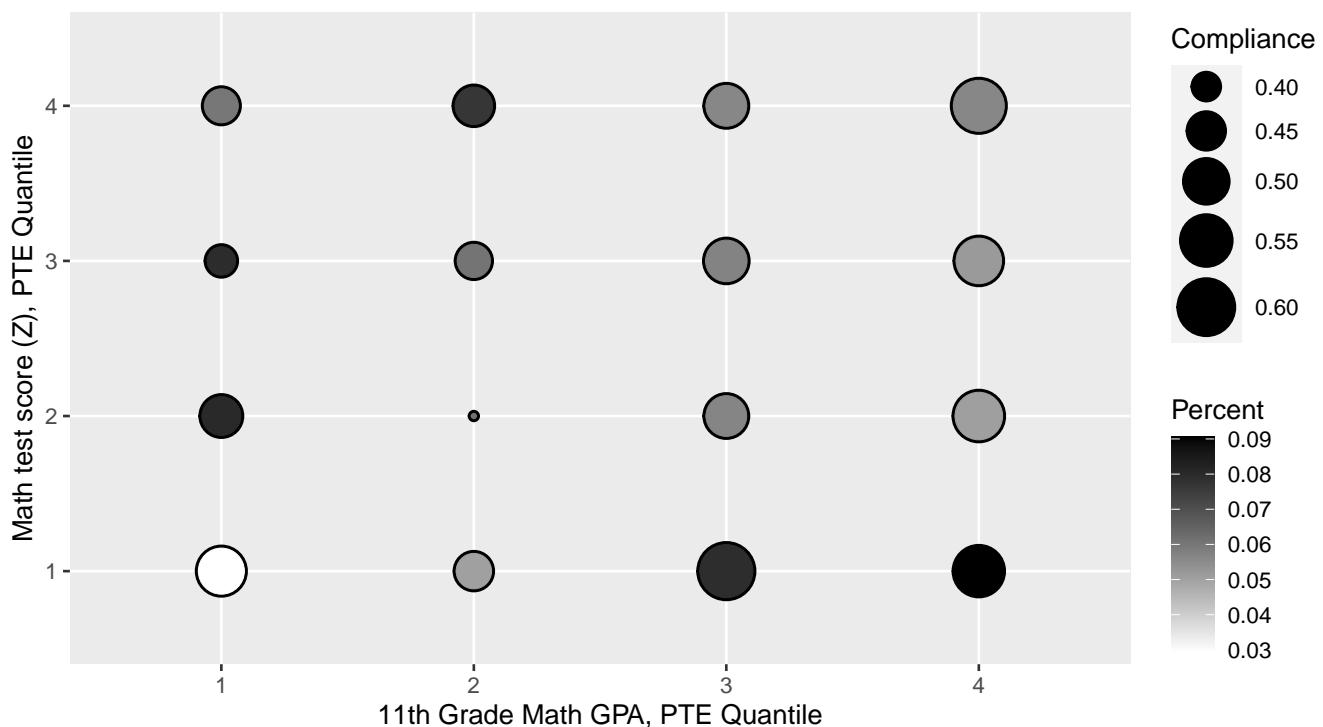
PTE Densities: Non–math GPA vs Graduated on–time



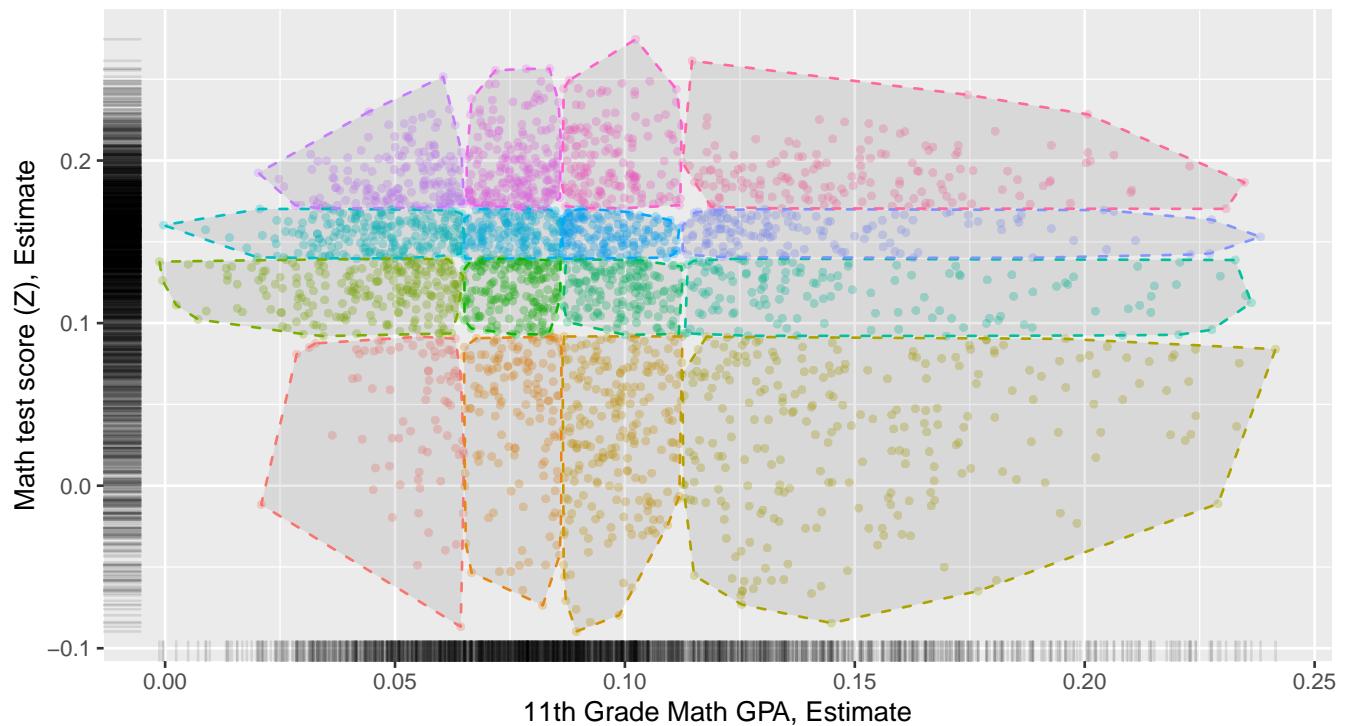
PTE Decile Plot: Math test score (Z) vs. 11th Grade Math GPA



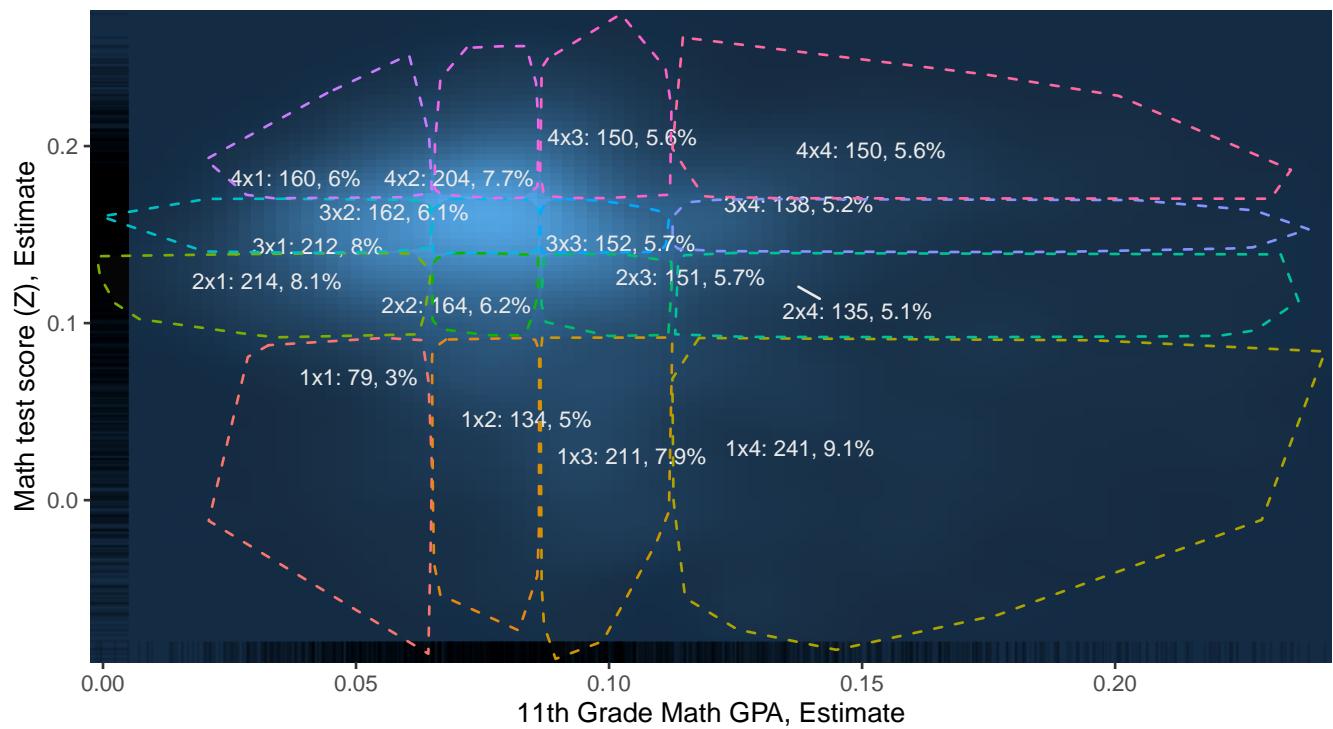
PTE Decile Plot: Math test score (Z) vs. 11th Grade Math GPA



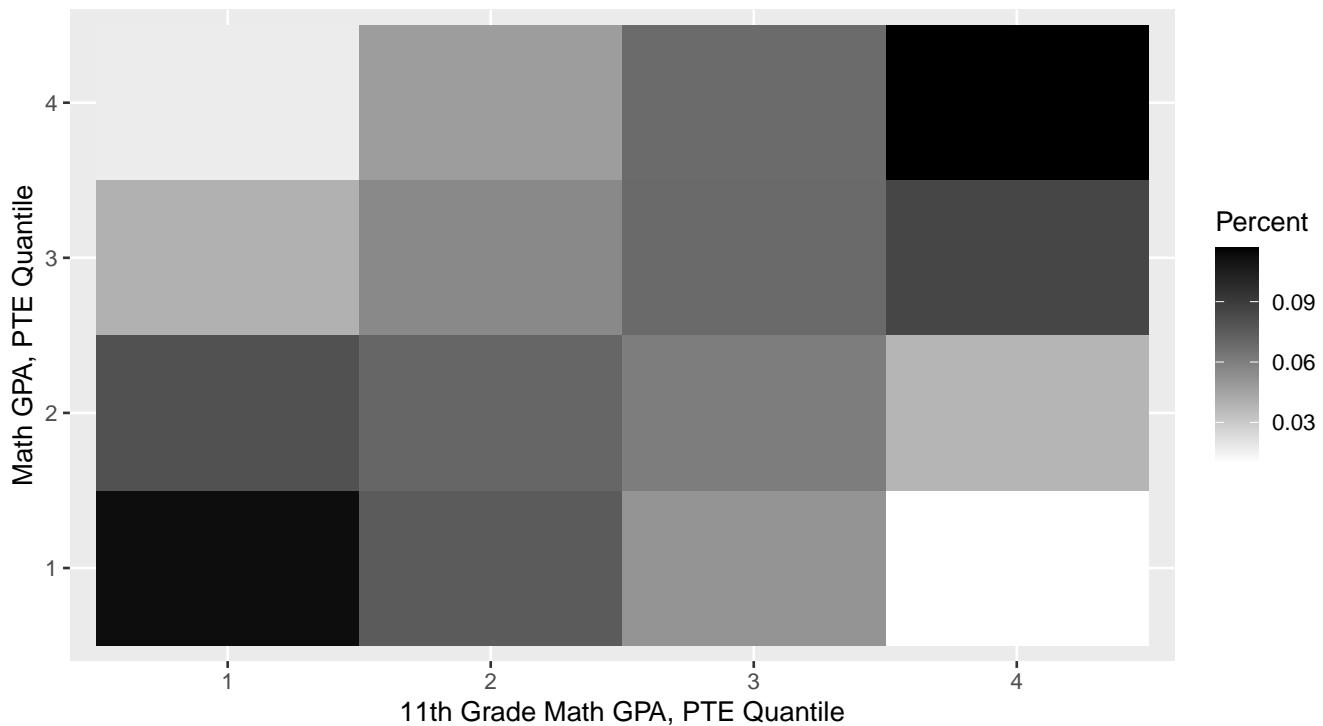
PTE Scatterplot: Math test score (Z) vs 11th Grade Math GPA



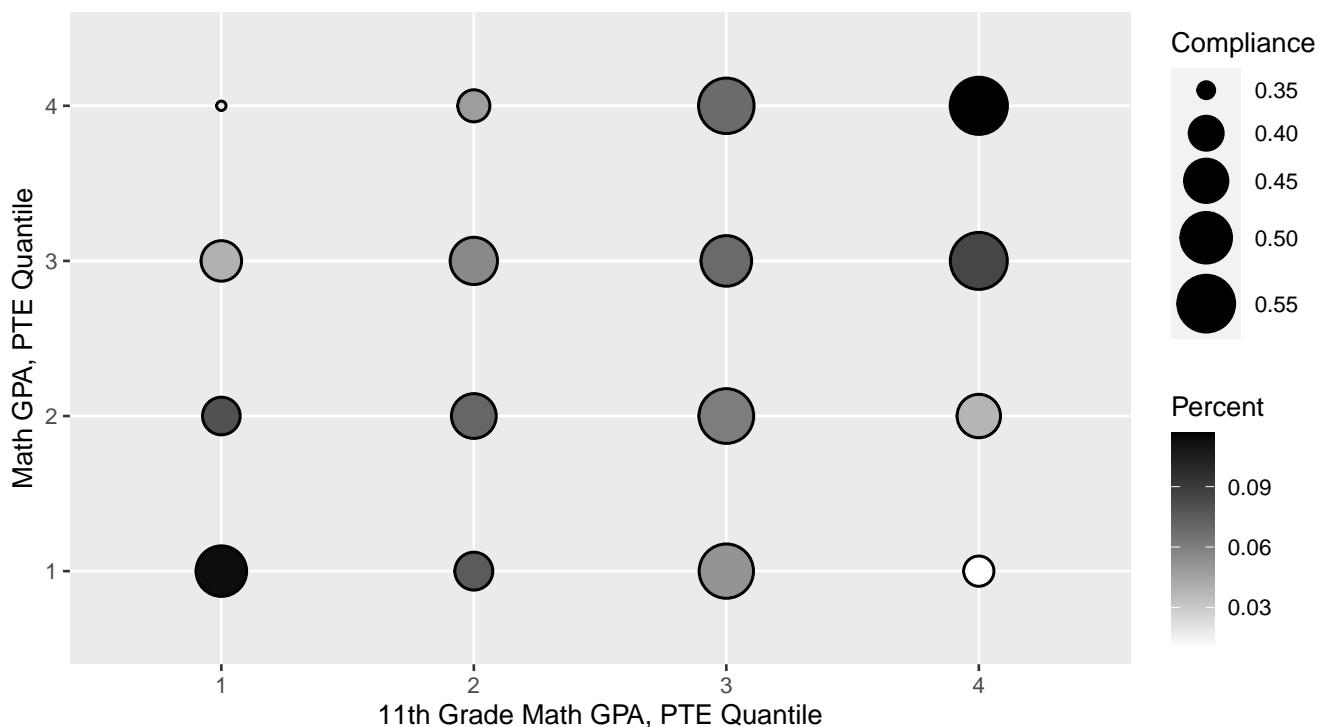
PTE Densities: Math test score (Z) vs 11th Grade Math GPA



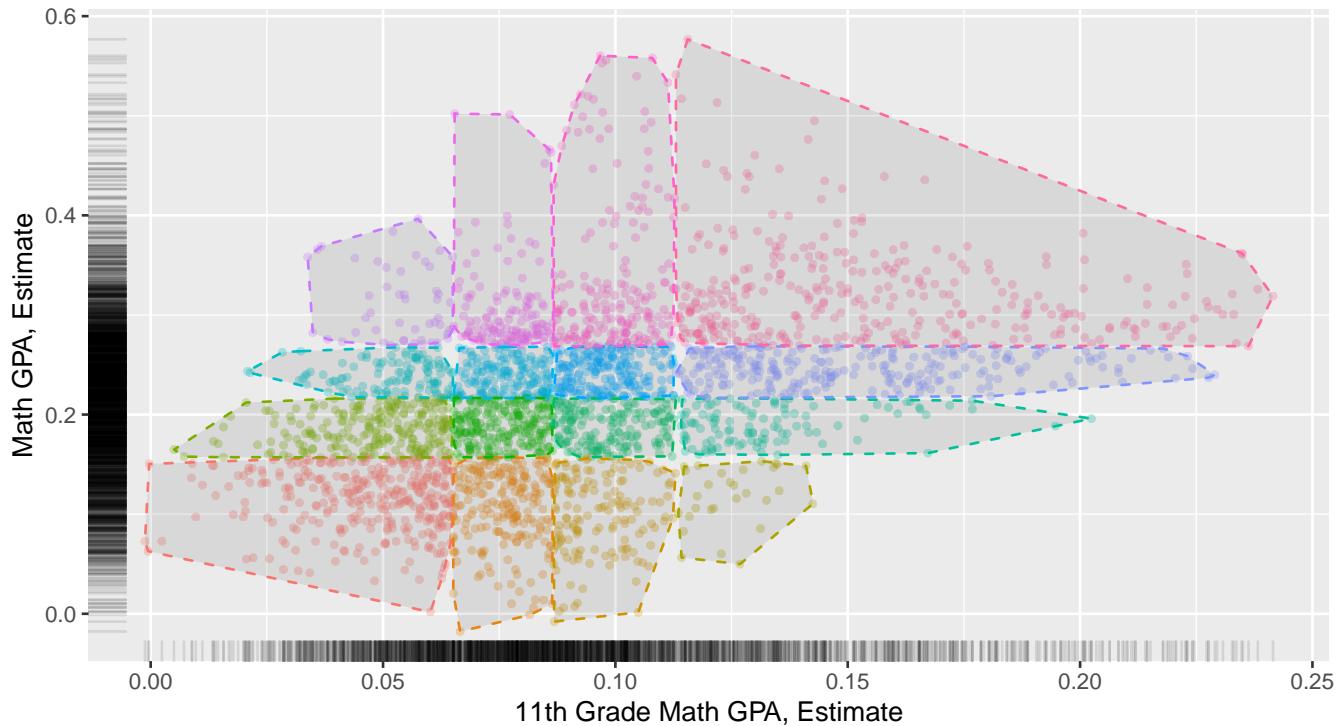
PTE Decile Plot: Math GPA vs. 11th Grade Math GPA



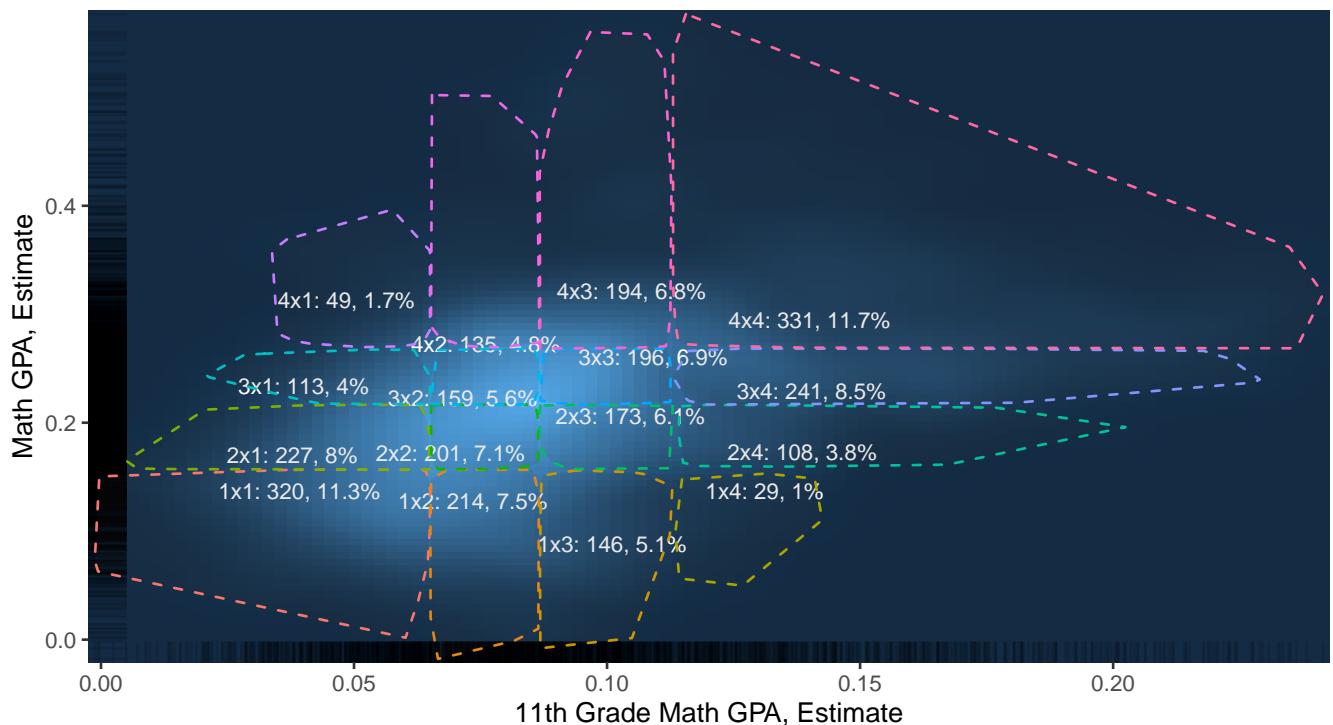
PTE Decile Plot: Math GPA vs. 11th Grade Math GPA



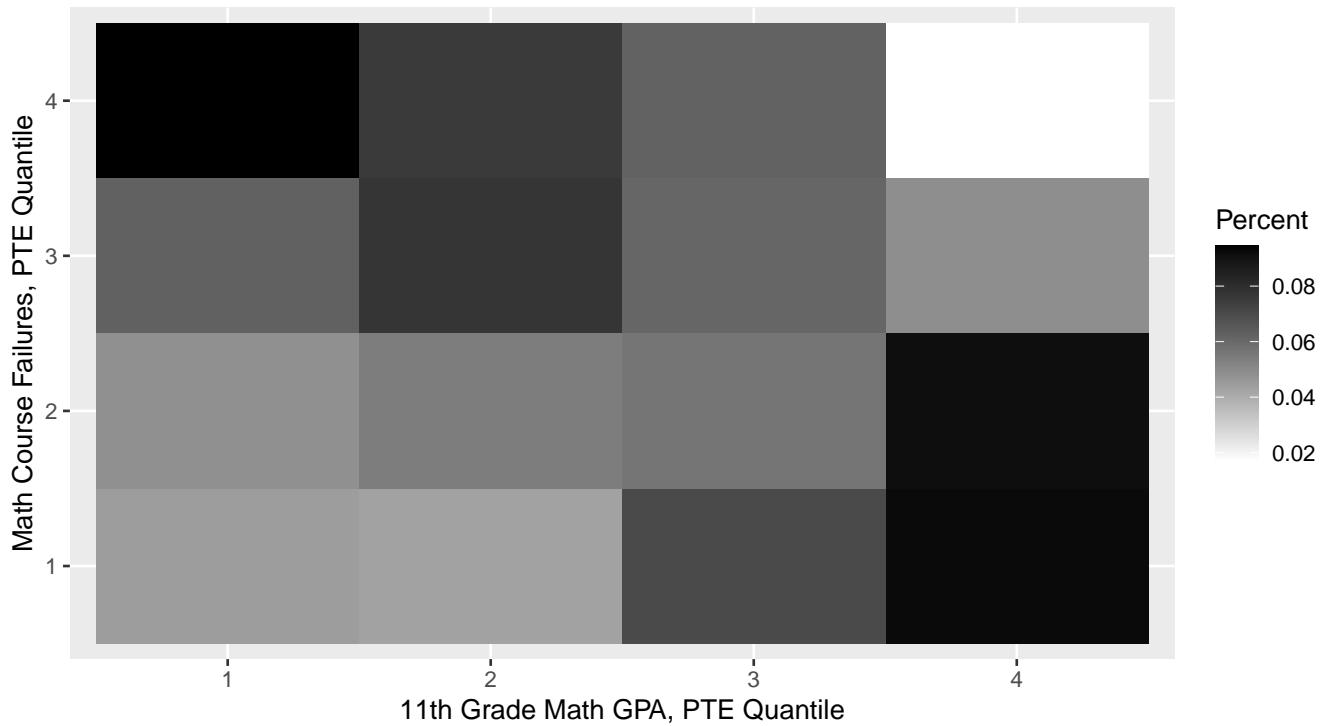
PTE Scatterplot: Math GPA vs 11th Grade Math GPA



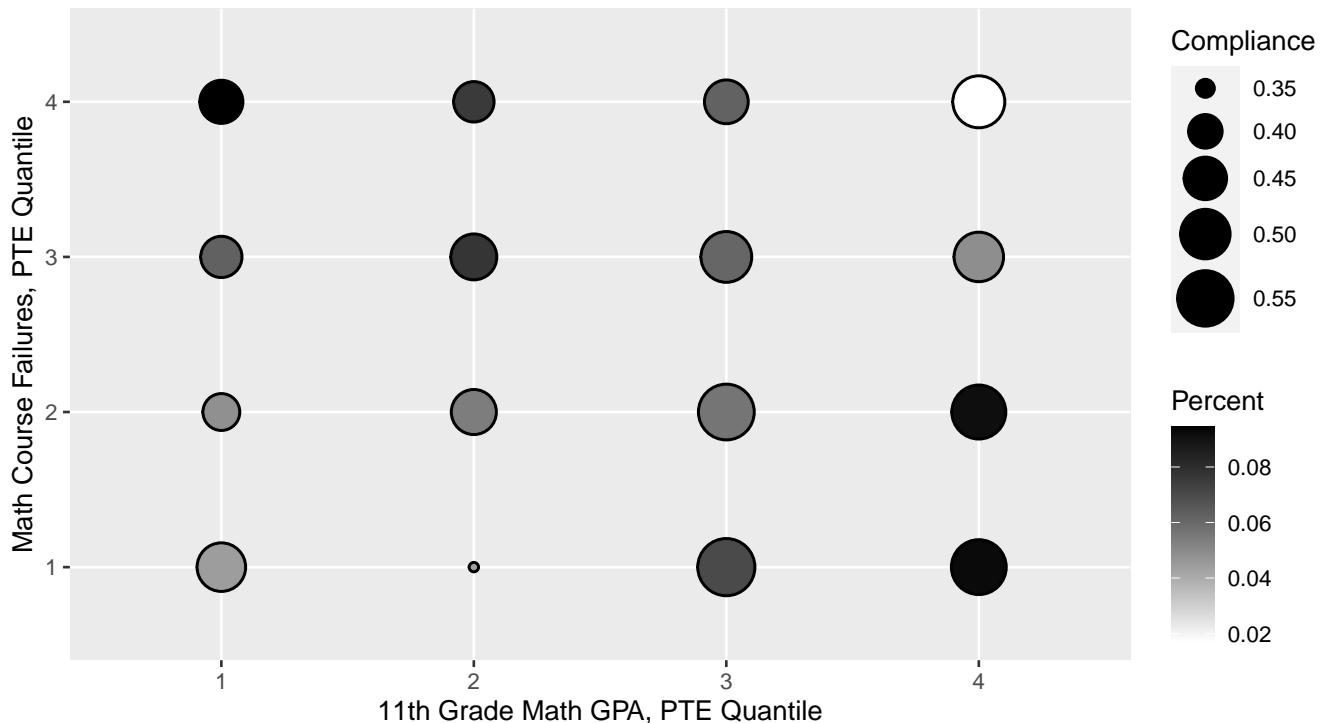
PTE Densities: Math GPA vs 11th Grade Math GPA



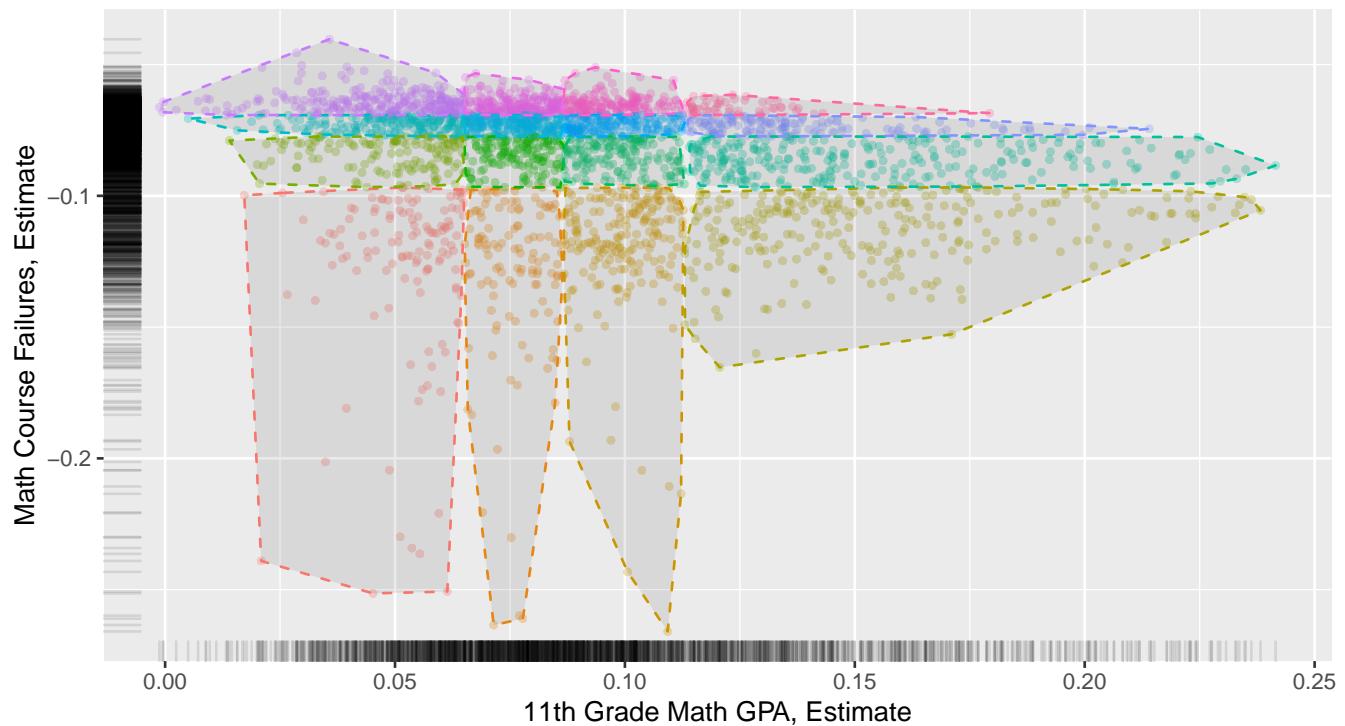
PTE Decile Plot: Math Course Failures vs. 11th Grade Math GPA



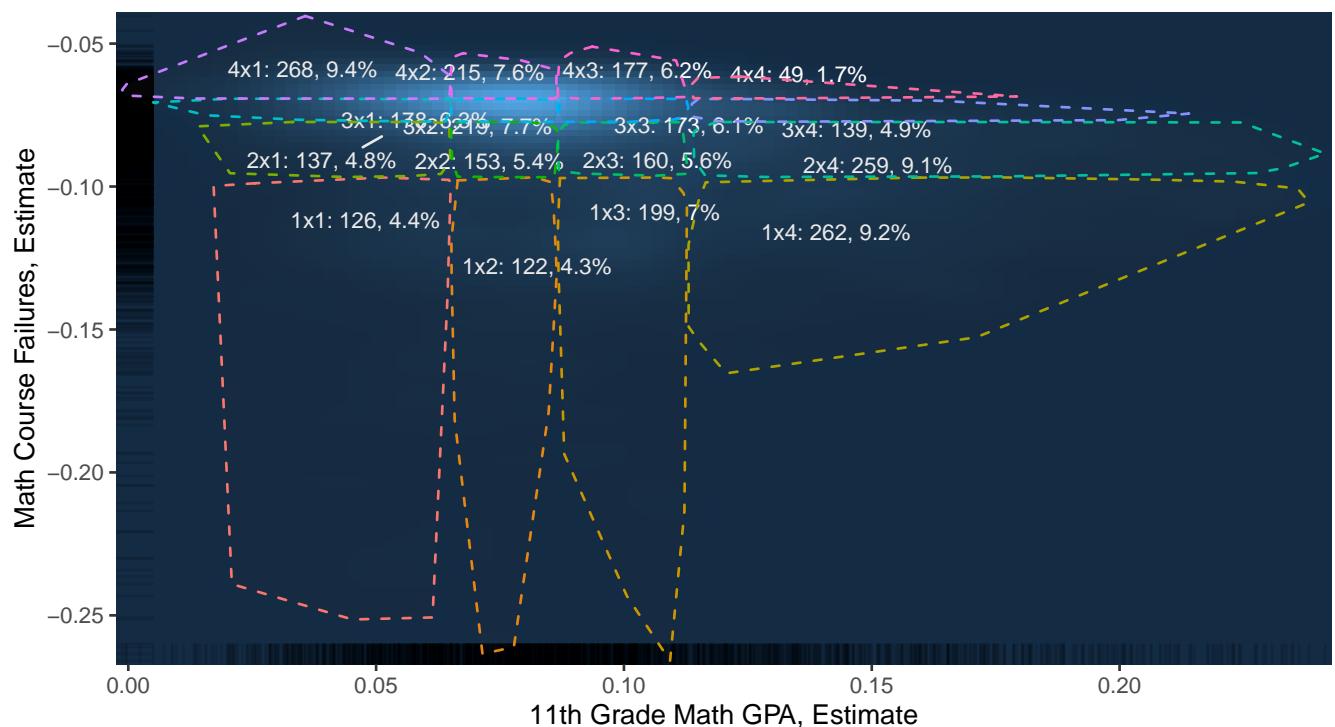
PTE Decile Plot: Math Course Failures vs. 11th Grade Math GPA



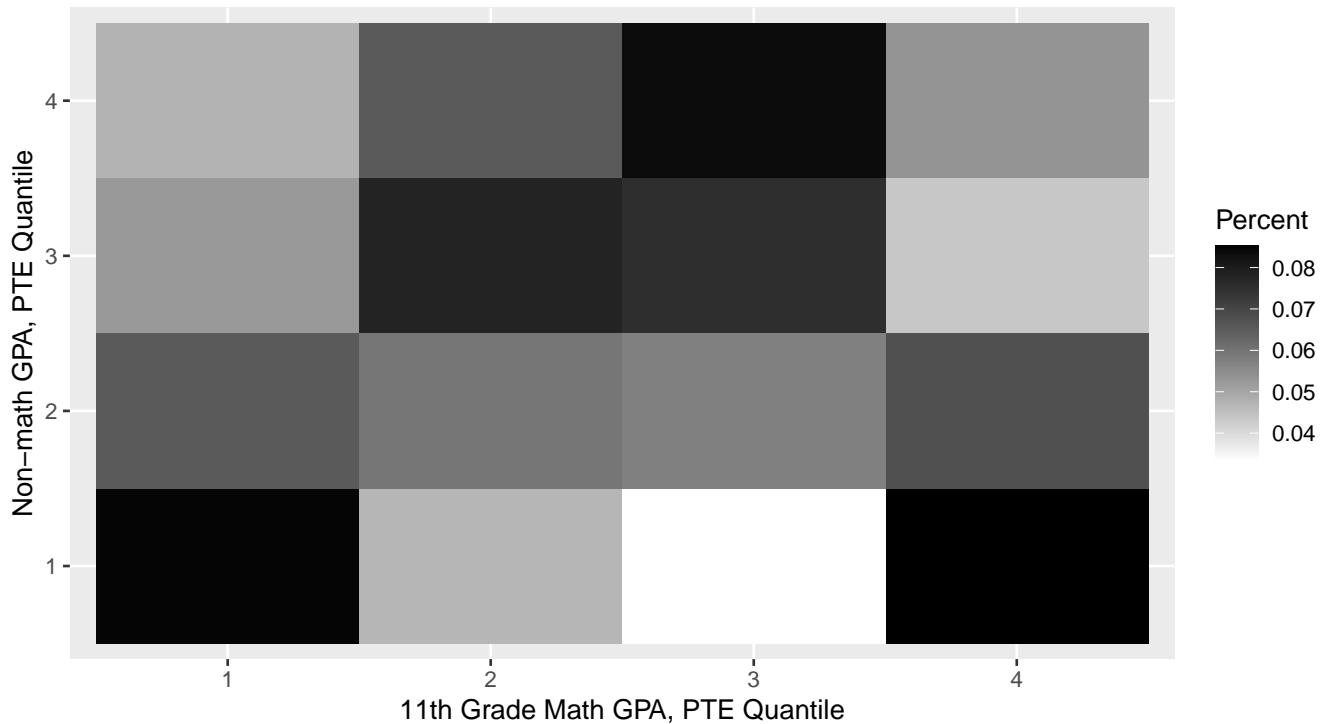
PTE Scatterplot: Math Course Failures vs 11th Grade Math GPA



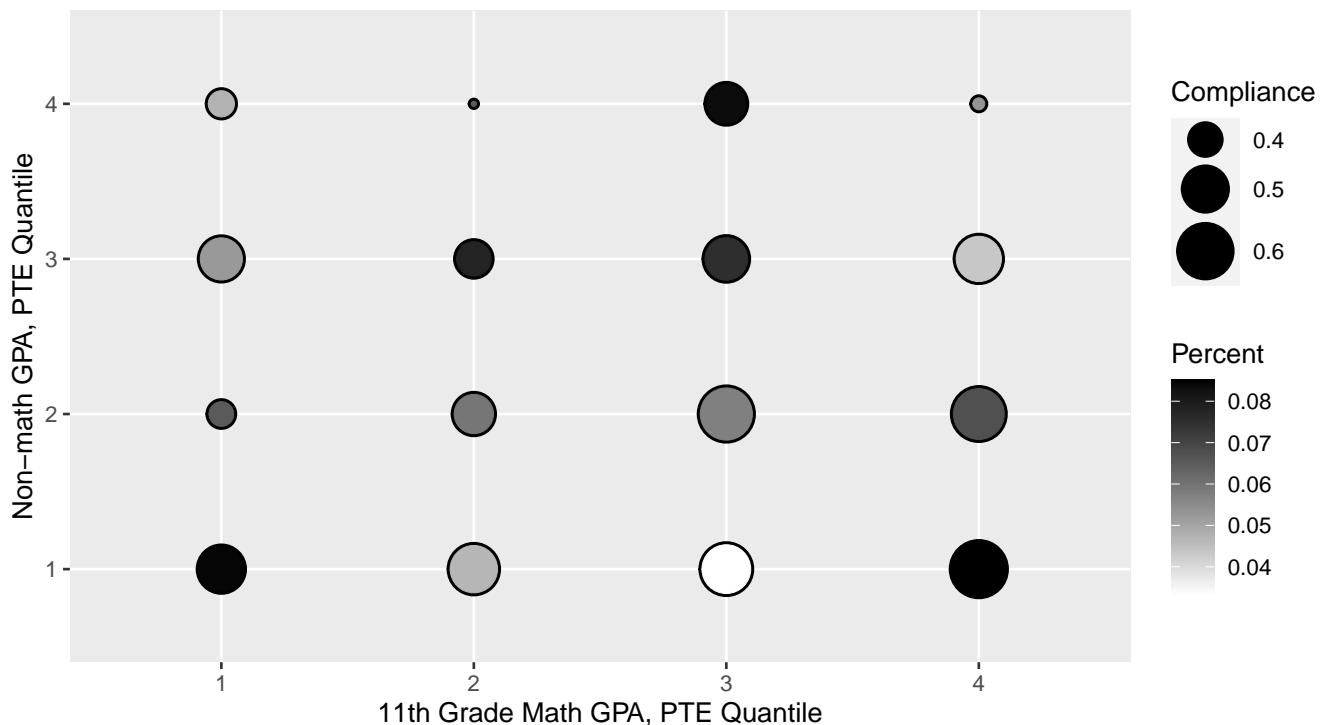
PTE Densities: Math Course Failures vs 11th Grade Math GPA



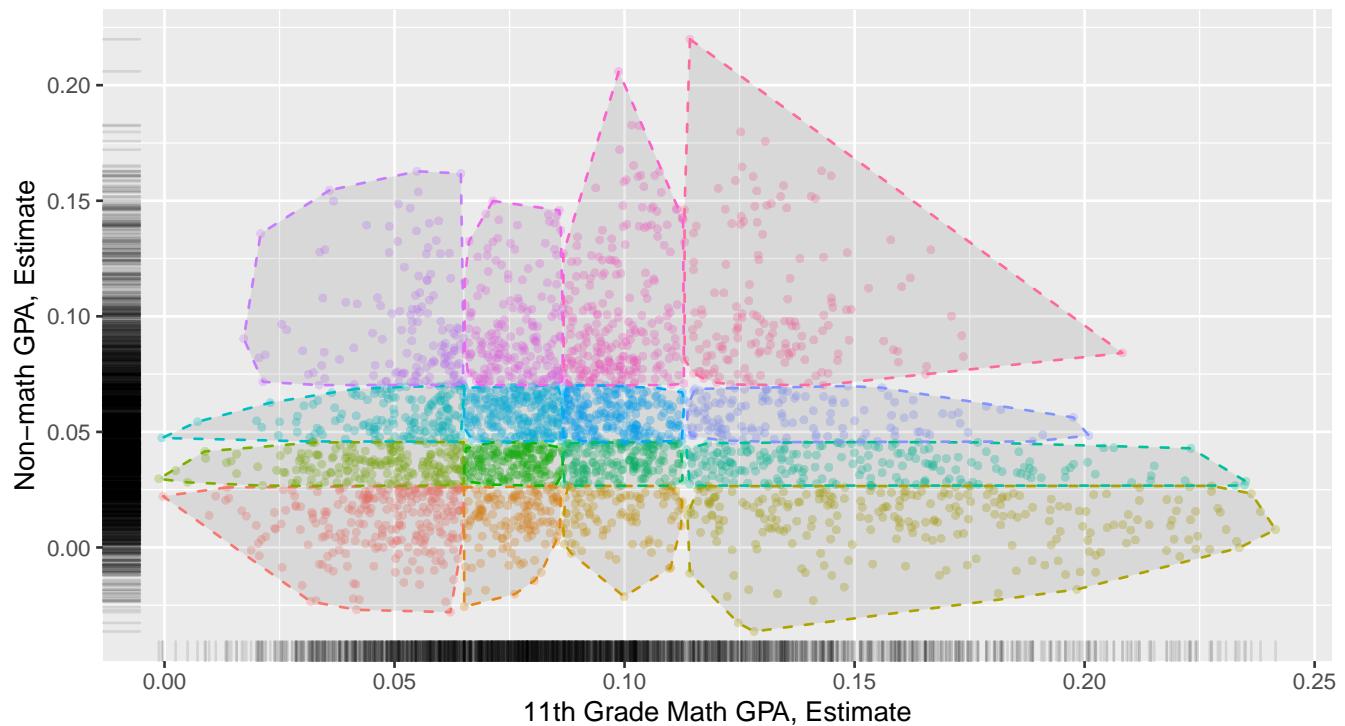
PTE Decile Plot: Non–math GPA vs. 11th Grade Math GPA



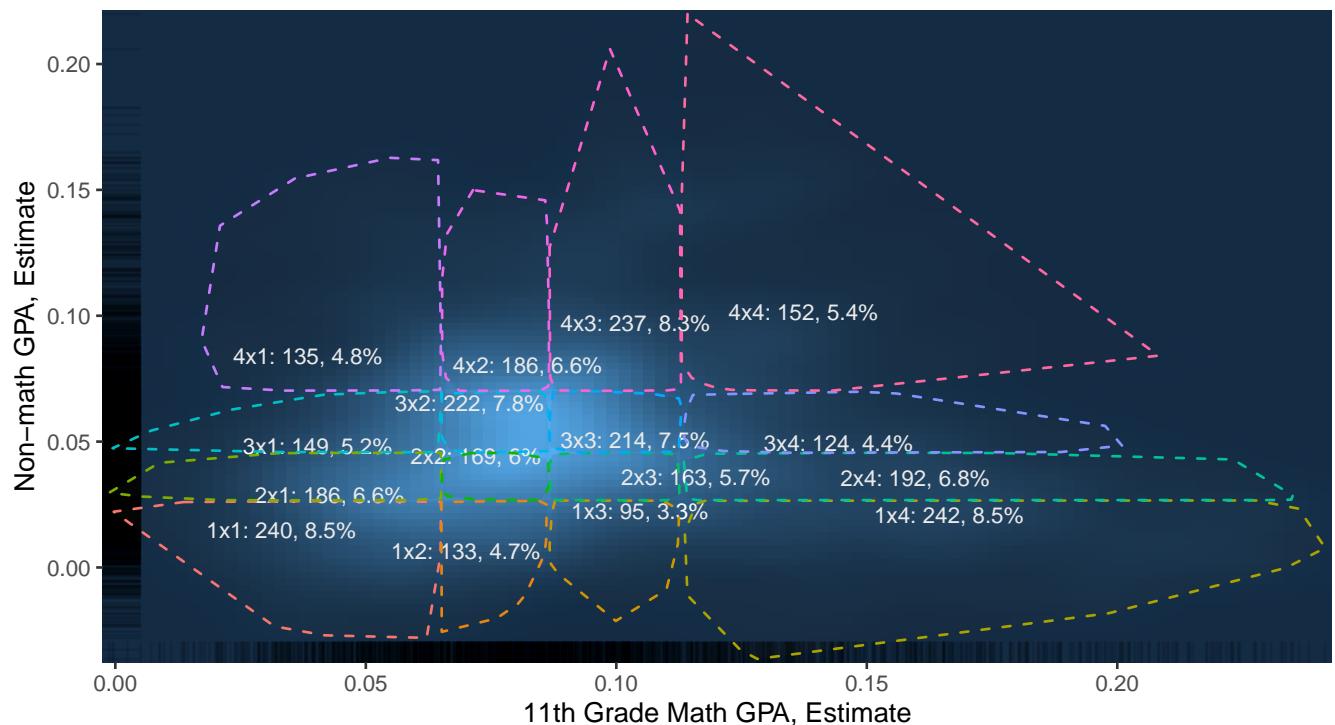
PTE Decile Plot: Non–math GPA vs. 11th Grade Math GPA



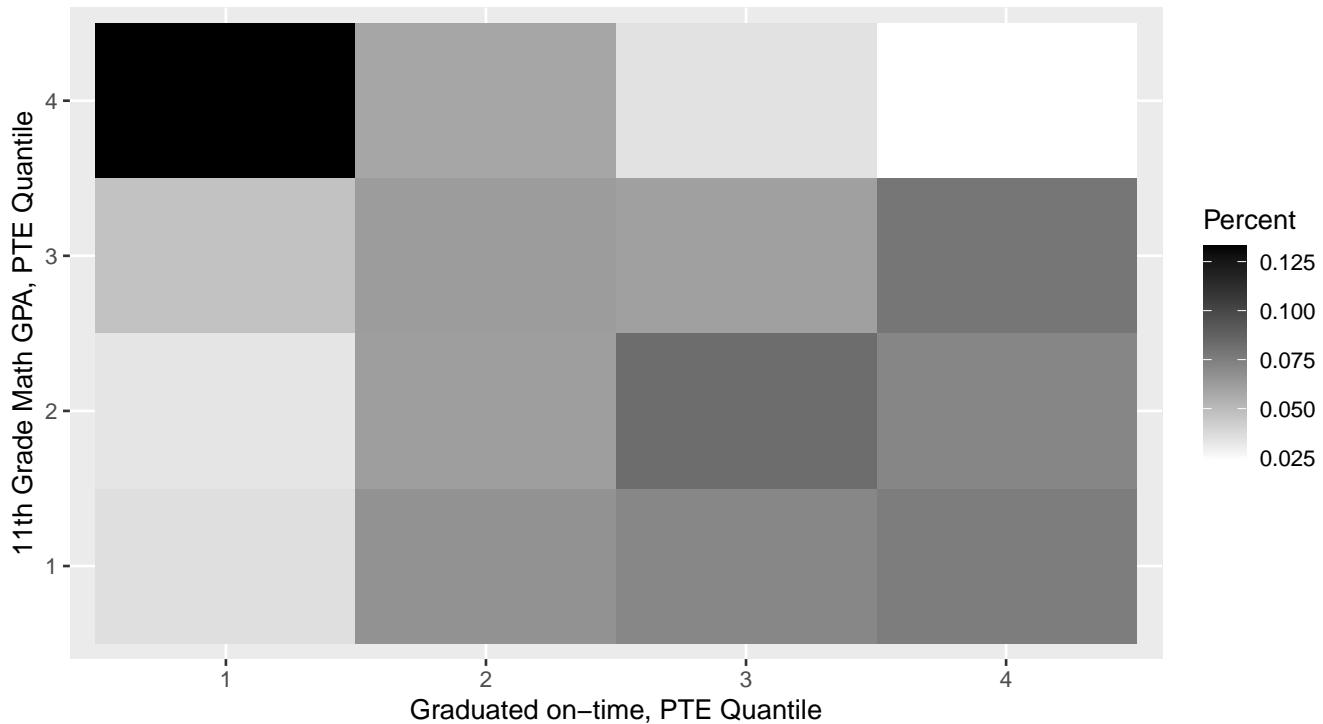
PTE Scatterplot: Non–math GPA vs 11th Grade Math GPA



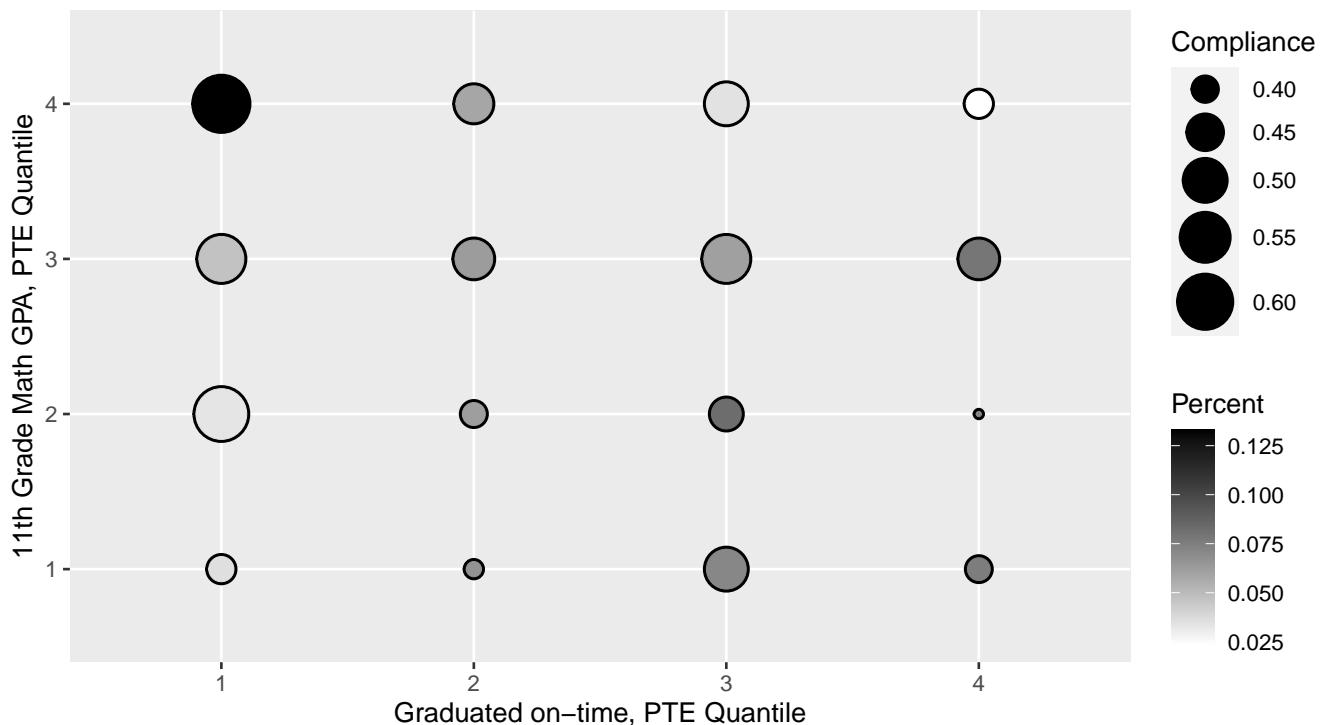
PTE Densities: Non–math GPA vs 11th Grade Math GPA



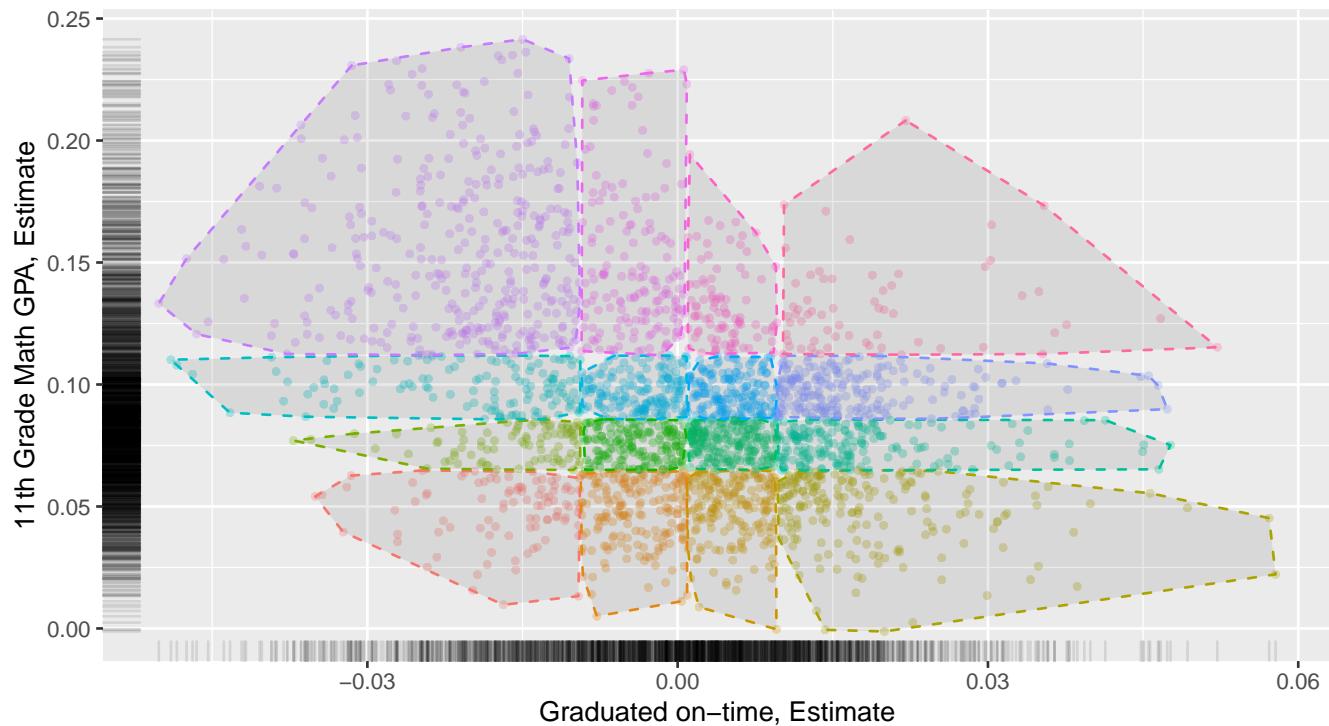
PTE Decile Plot: 11th Grade Math GPA vs. Graduated on-time



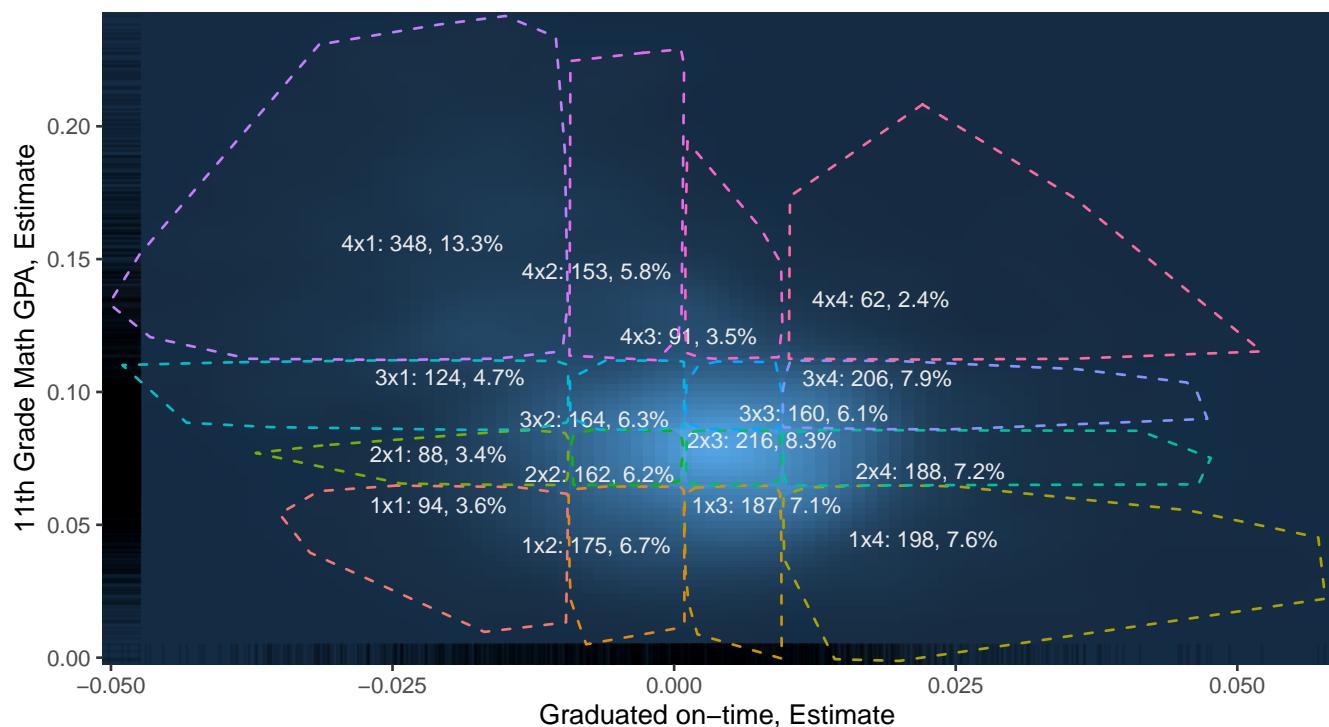
PTE Decile Plot: 11th Grade Math GPA vs. Graduated on-time



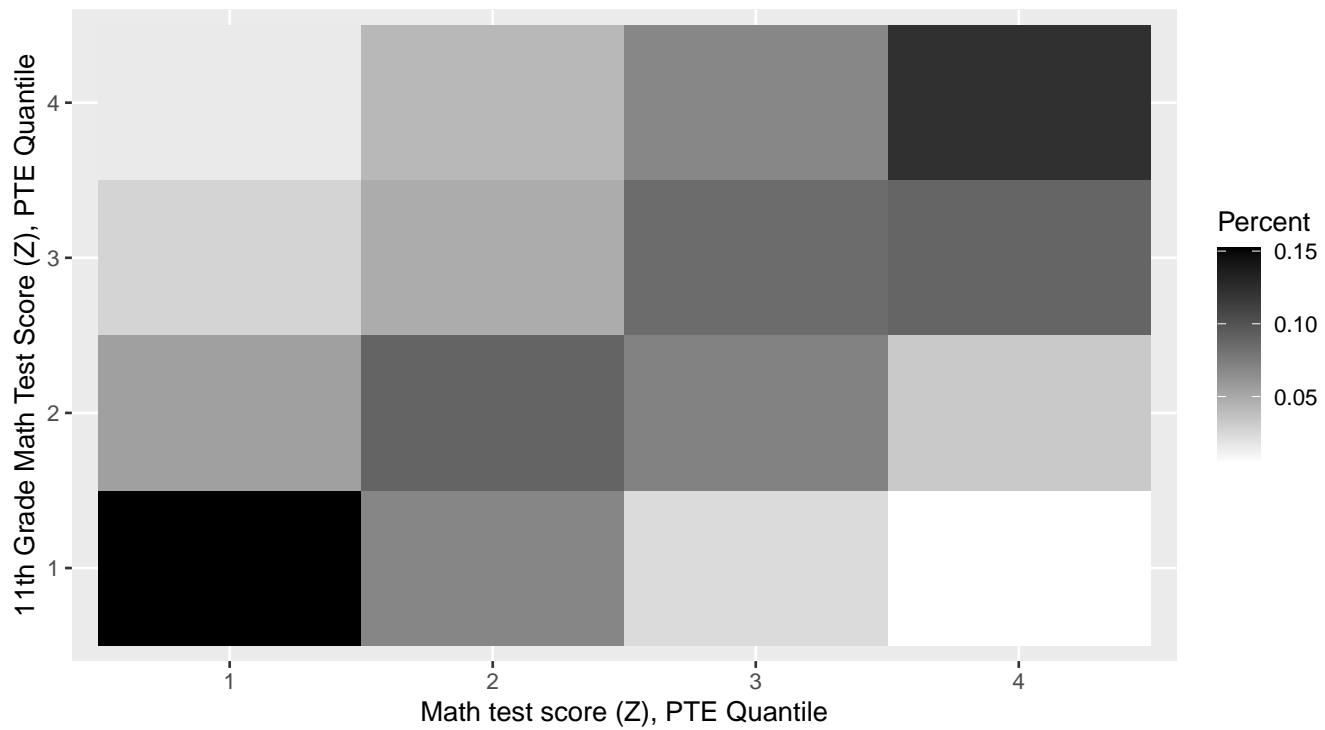
PTE Scatterplot: 11th Grade Math GPA vs Graduated on-time



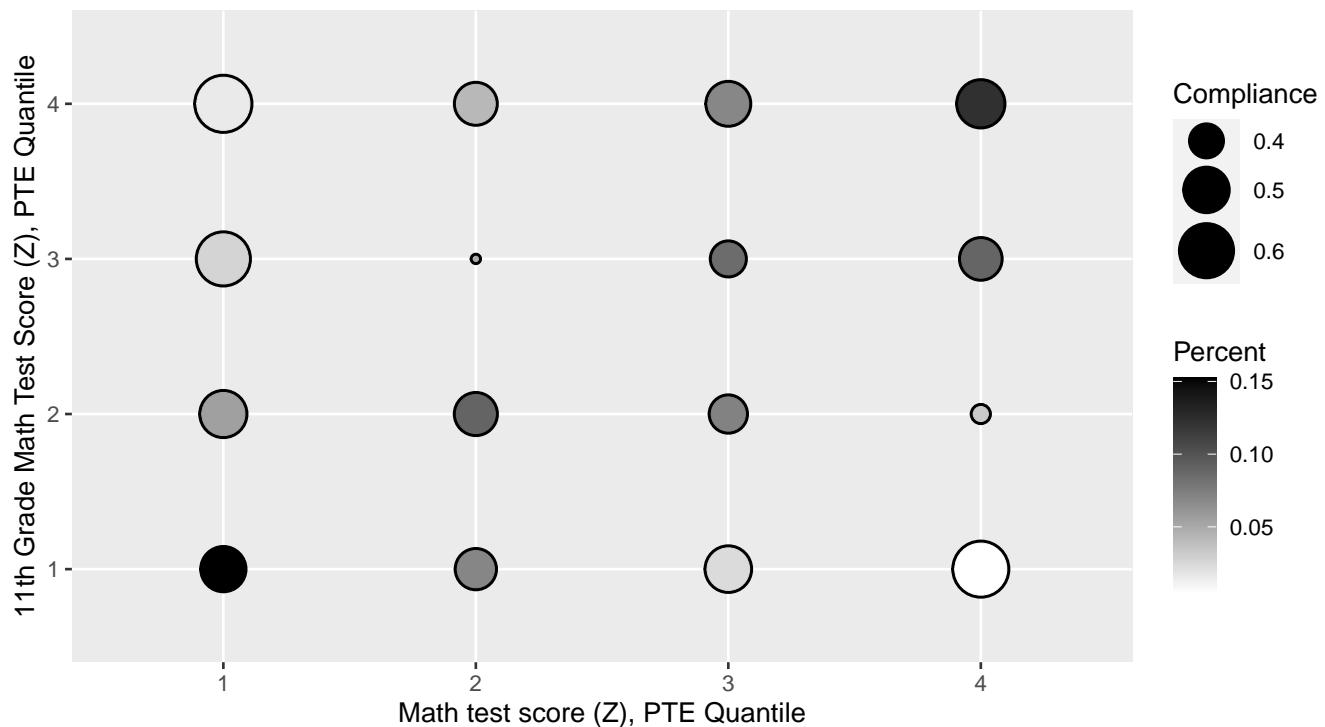
PTE Densities: 11th Grade Math GPA vs Graduated on-time



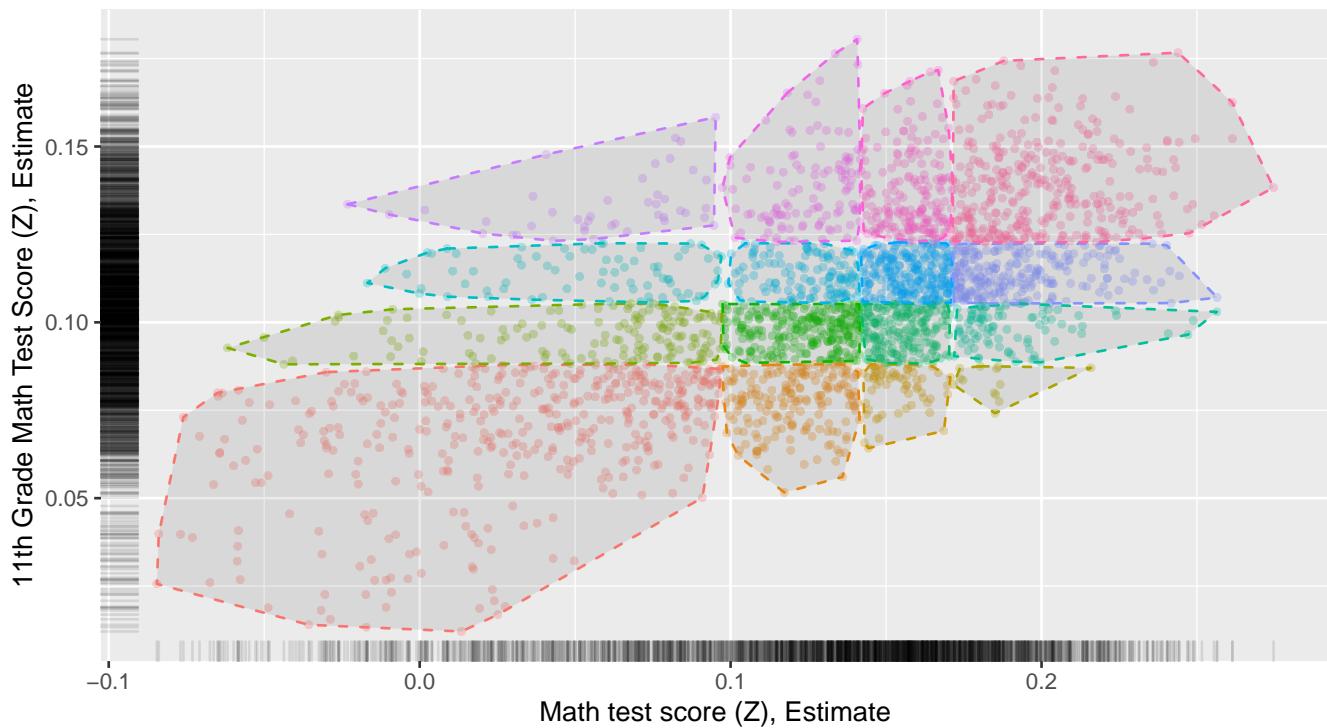
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. Math test score (Z)



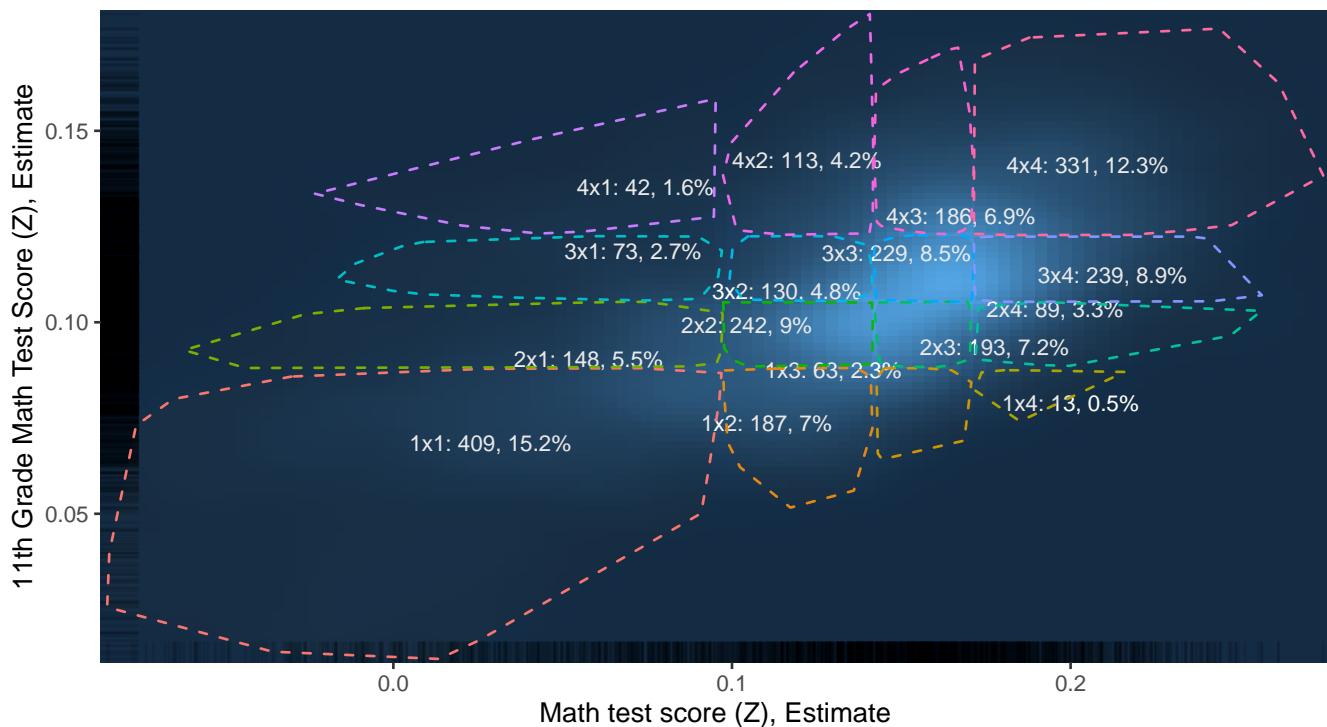
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. Math test score (Z)



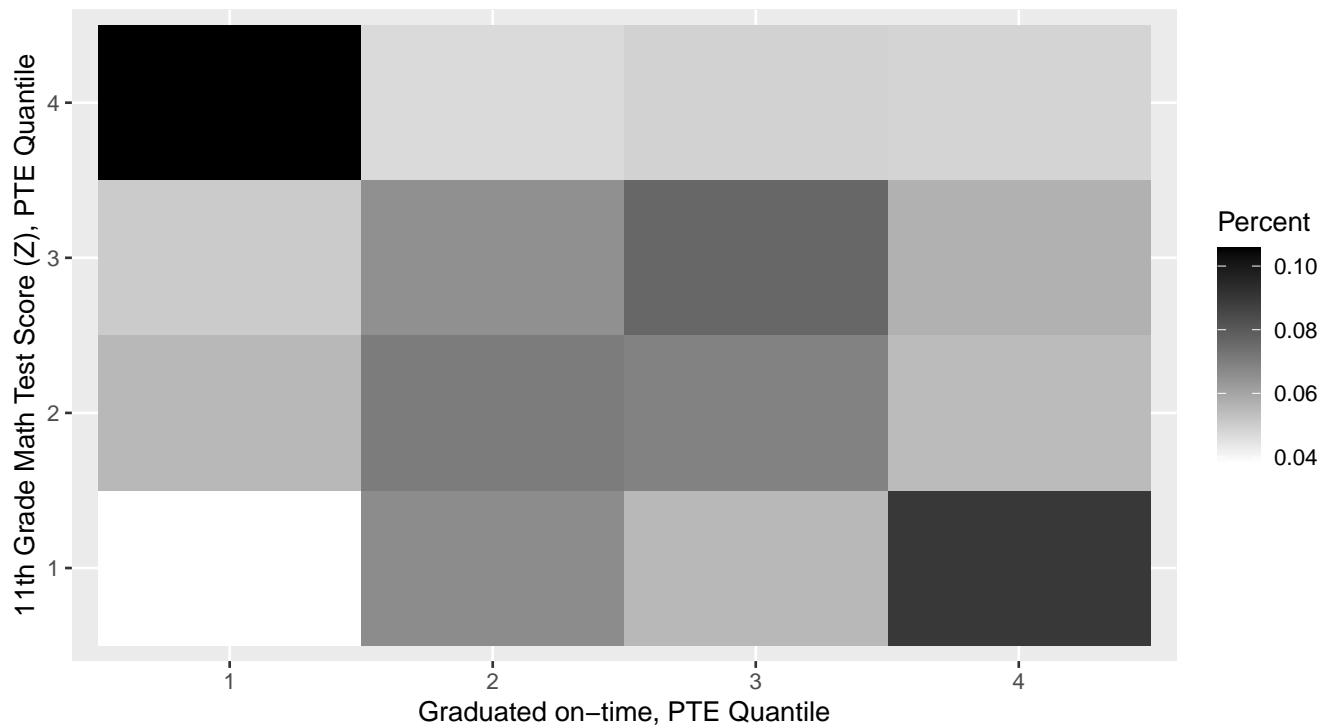
PTE Scatterplot: 11th Grade Math Test Score (Z) vs Math test score (Z)



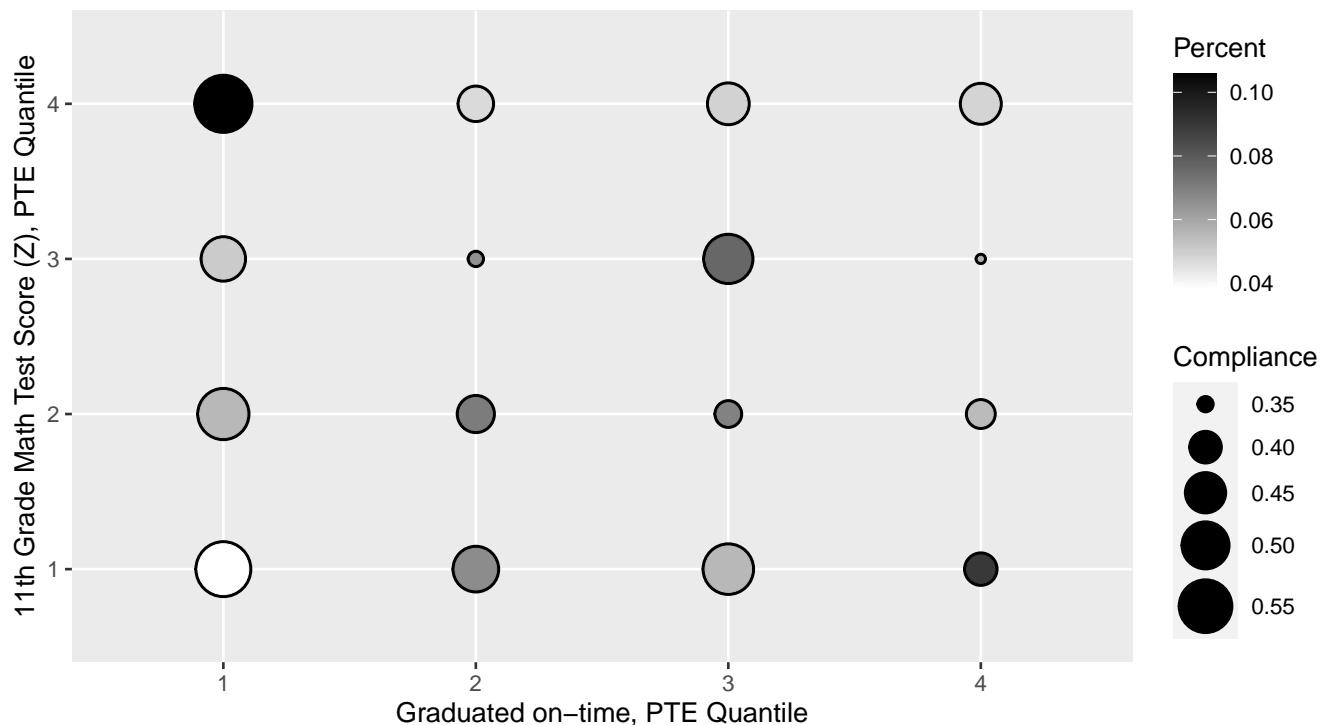
PTE Densities: 11th Grade Math Test Score (Z) vs Math test score (Z)



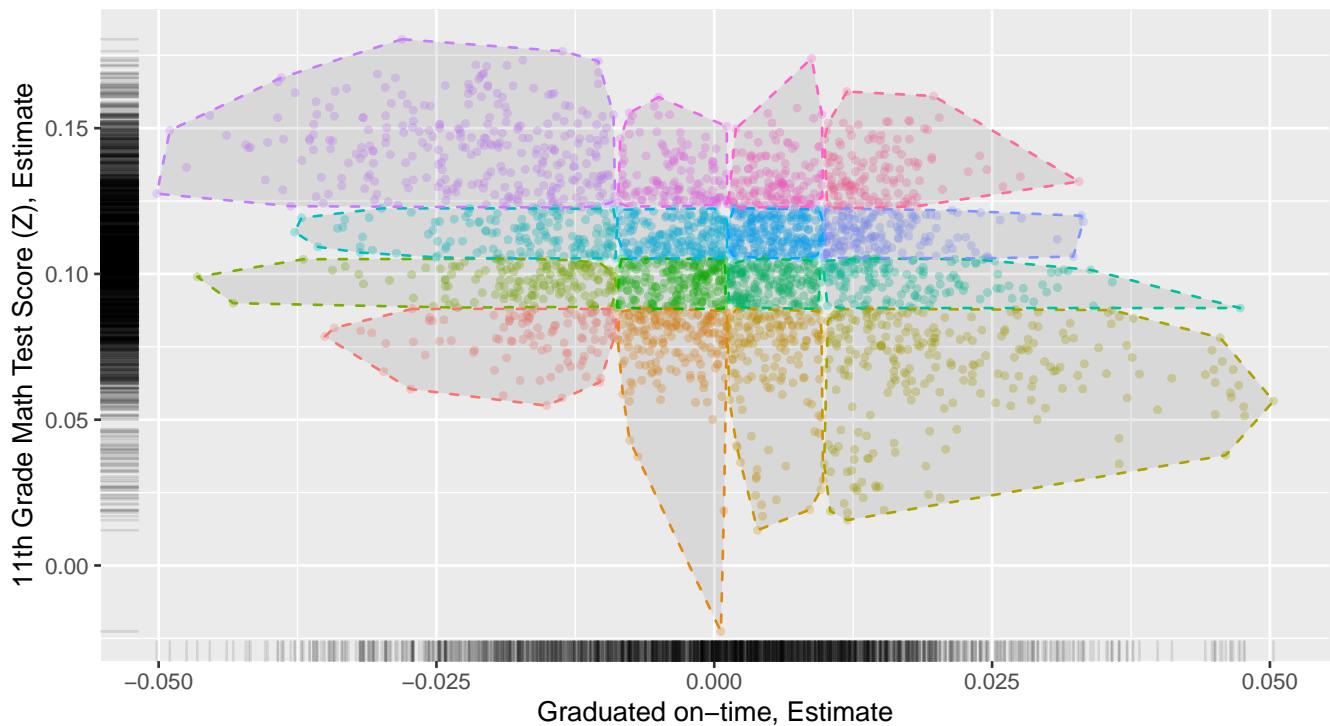
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. Graduated on-time



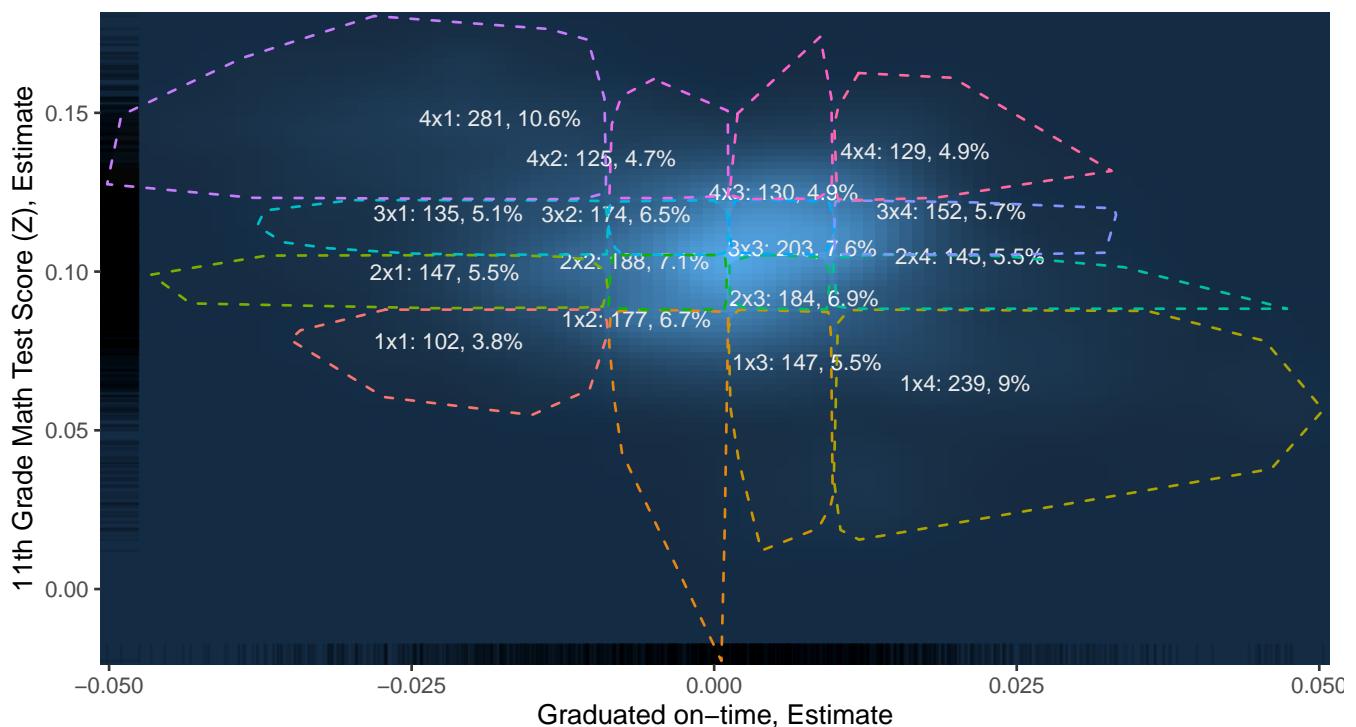
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. Graduated on-time



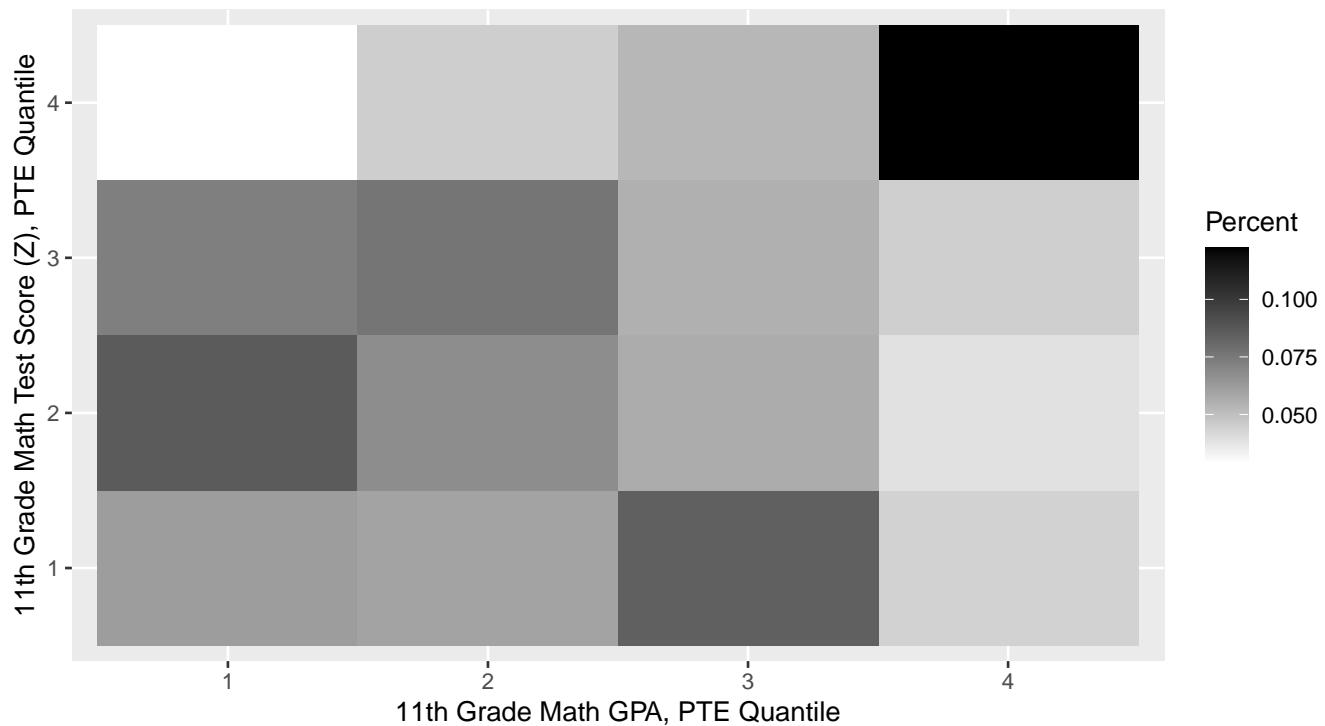
PTE Scatterplot: 11th Grade Math Test Score (Z) vs Graduated on-time



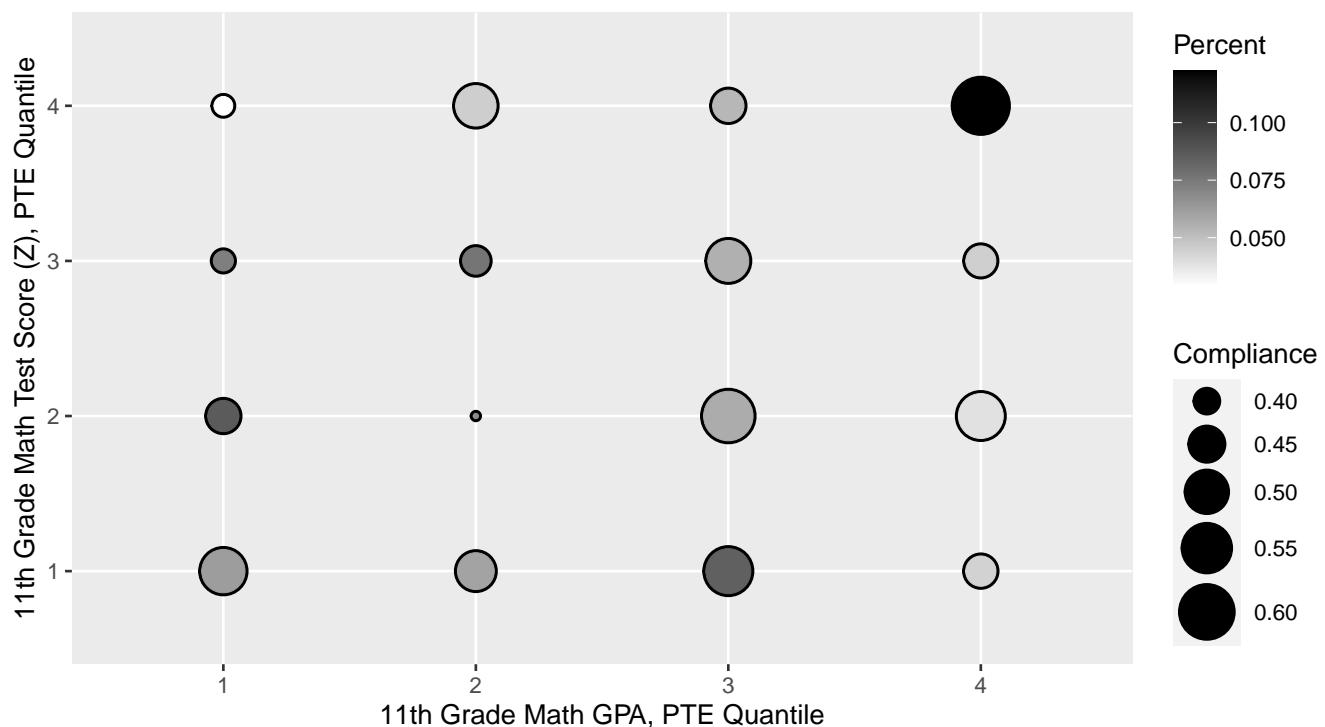
PTE Densities: 11th Grade Math Test Score (Z) vs Graduated on-time



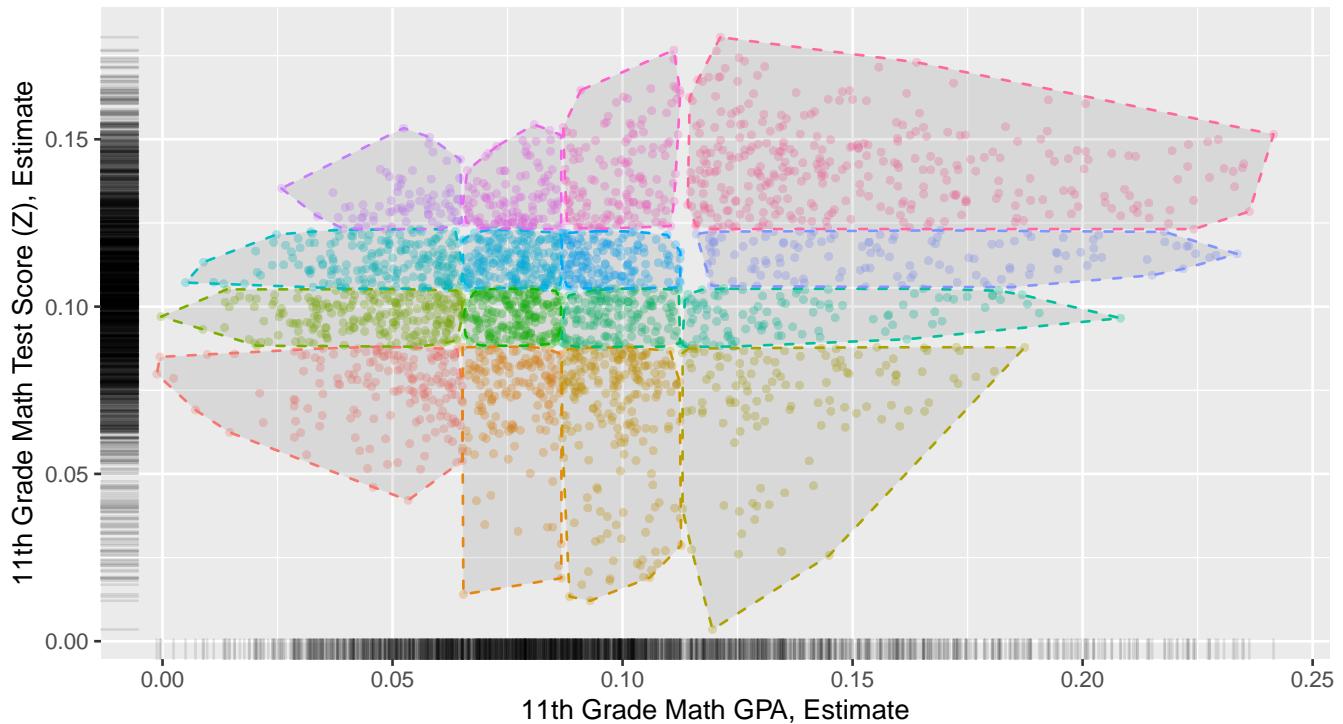
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. 11th Grade Math GPA



PTE Decile Plot: 11th Grade Math Test Score (Z) vs. 11th Grade Math GPA



PTE Scatterplot: 11th Grade Math Test Score (Z) vs 11th Grade Math GPA



PTE Densities: 11th Grade Math Test Score (Z) vs 11th Grade Math GPA

