

# 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:

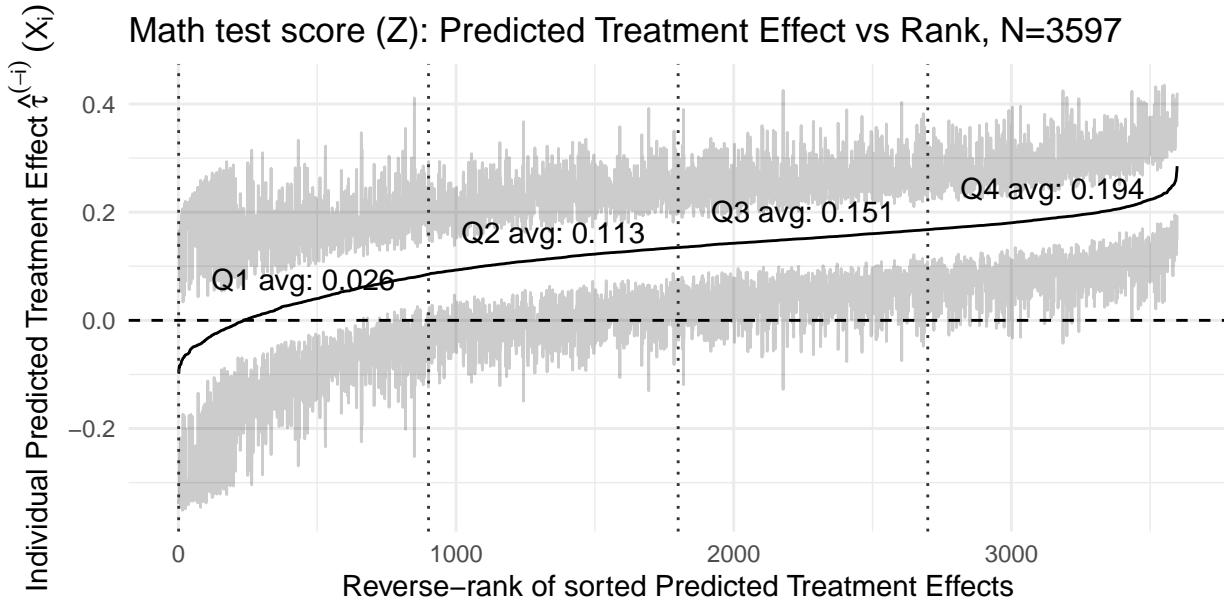
- Math Test Scores
- Math Class Failures
- Math GPA
- Reading Test Scores
- Non-math GPA
- Graduated on-time
- Ever graduated
- Participation in Saga tutoring in year 2
- 11th Grade Math GPA

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.

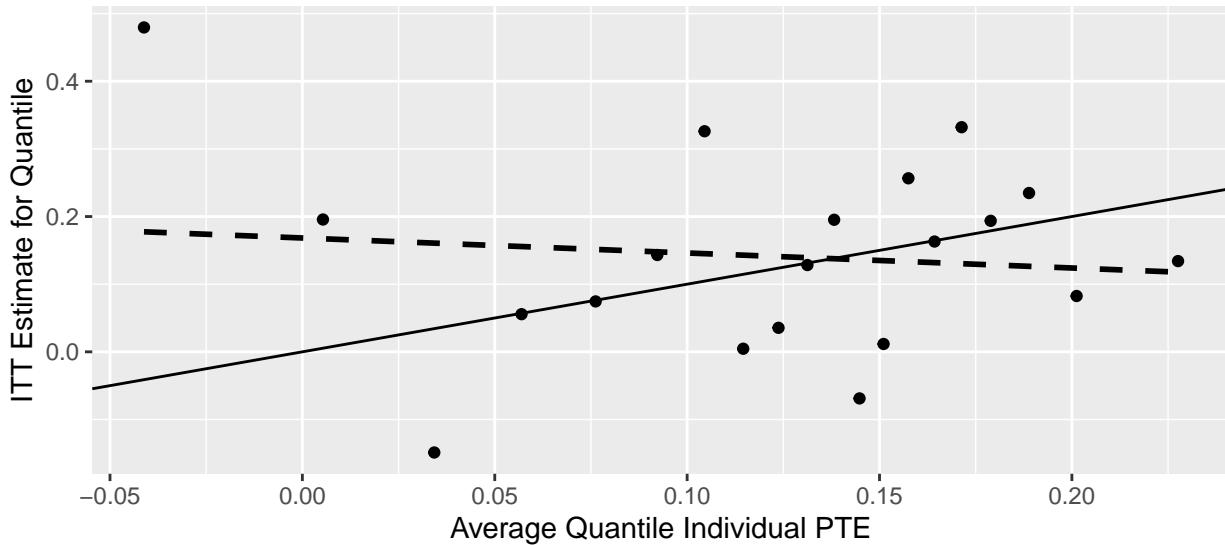
We use all covariates from our main analyses (except for randomization block). These include gender, age, learning disability, free lunch recipient, race, baseline grade level, GPA, baseline test performance (and within-baseline-school math test decile), days absent from school, disciplinary incidents,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

## Outcome: Math test score (Z)



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 0.121 (0.026). We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.003 and a ‘differential forest prediction’ (DFP) coefficient of 0.325, with a corresponding DFP p-value of 0.223.<sup>1</sup>

## Calibration Plot: Math test score (Z)



<sup>1</sup>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	0.121 +/- 0.051	0.026
Individual PTE Quartile 4	0.143 +/- 0.091	0.047
Individual PTE Quartile 3	0.139 +/- 0.1	0.051
Individual PTE Quartile 2	0.174 +/- 0.099	0.050
Individual PTE Quartile 1	0.031 +/- 0.113	0.058
Top 2 PTE Quartiles	0.141 +/- 0.068	0.034
Bottom 2 PTE Quartiles	0.102 +/- 0.075	0.038
Bottom 3 PTE Quartiles	0.114 +/- 0.06	0.031

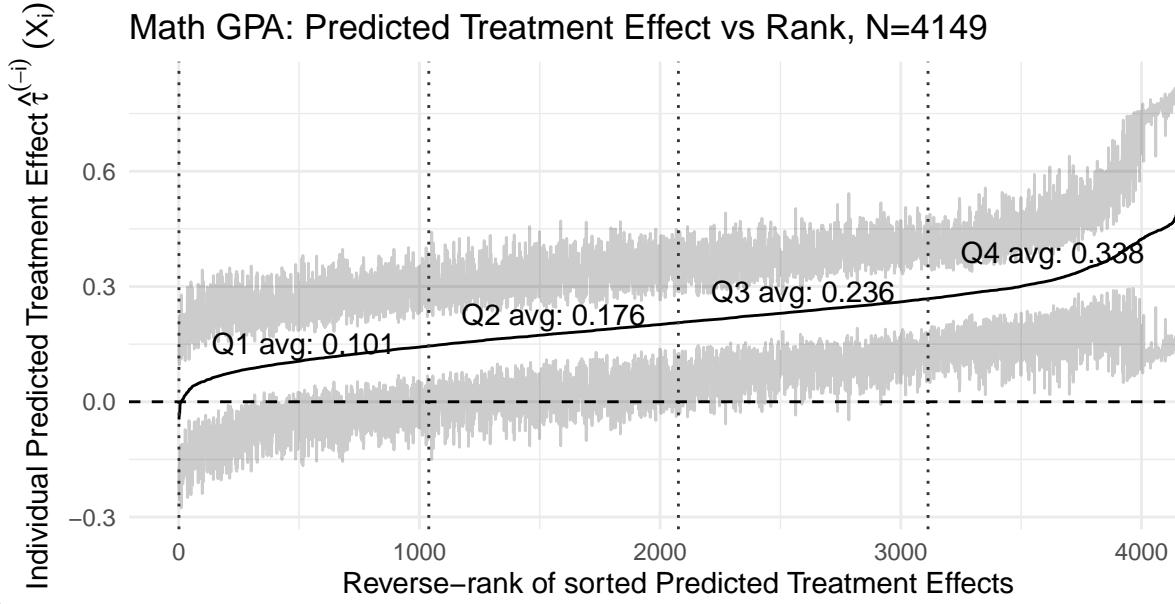
Table 2: Differences between subsample average treatment effects: Math test score (Z)

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.08, 0.138]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.062, 0.14]
Quartile 4	Quartile 1	[-0.033, 0.257]

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 $\hat{\tau}$	0.026	0.113	0.151	0.194
N	900.000	899.000	899.000	899.000
Age	14.731	14.661	14.476	14.320
Female	0.153	0.130	0.172	0.180
Has IEP	0.368	0.210	0.080	0.021
Has Free/Reduced Lunch	0.892	0.898	0.884	0.898
Black	0.592	0.536	0.531	0.380
Hispanic	0.373	0.434	0.424	0.554
Other Race	0.034	0.030	0.046	0.066
In 9th Grade	0.742	0.655	0.726	0.791
In 10th Grade	0.252	0.344	0.274	0.209
Baseline GPA	1.811	2.408	2.632	2.622
Num. A's	3.370	6.560	8.870	10.591
Num. B's	4.208	7.903	8.784	8.929
Num. C's	7.713	8.561	7.987	8.684
Num. D's	4.487	2.998	2.496	3.107
Num. F's	4.227	2.050	1.532	1.681
Missing Baseline GPA/Grades	0.164	0.023	0.041	0.047
Days Absent	21.934	13.916	11.361	8.706
Missing Attendance Data	0.133	0.000	0.000	0.000
Math Test Score (Z)	-1.235	-0.034	0.399	0.537
Reading Test Score (Z)	-0.927	0.124	0.330	0.274
Missing Math Test	0.309	0.012	0.007	0.001
Missing Reading Test	0.310	0.018	0.009	0.003
Out-of-School Suspensions	2.247	0.846	0.515	0.130
Disciplinary Incidents	1.676	0.768	0.416	0.135
Any Arrests at Baseline	0.174	0.112	0.083	0.041
Arrests: Violent Crime	0.119	0.049	0.051	0.013
Arrests: Property Crime	0.052	0.032	0.022	0.011
Arrests: Drug Crime	0.043	0.019	0.010	0.004
Math Score - Decile in Previous School	2.416	5.568	7.025	7.371
In Study 2	0.459	0.442	0.488	0.568
Participated in Year 1 of Study	0.296	0.260	0.241	0.273

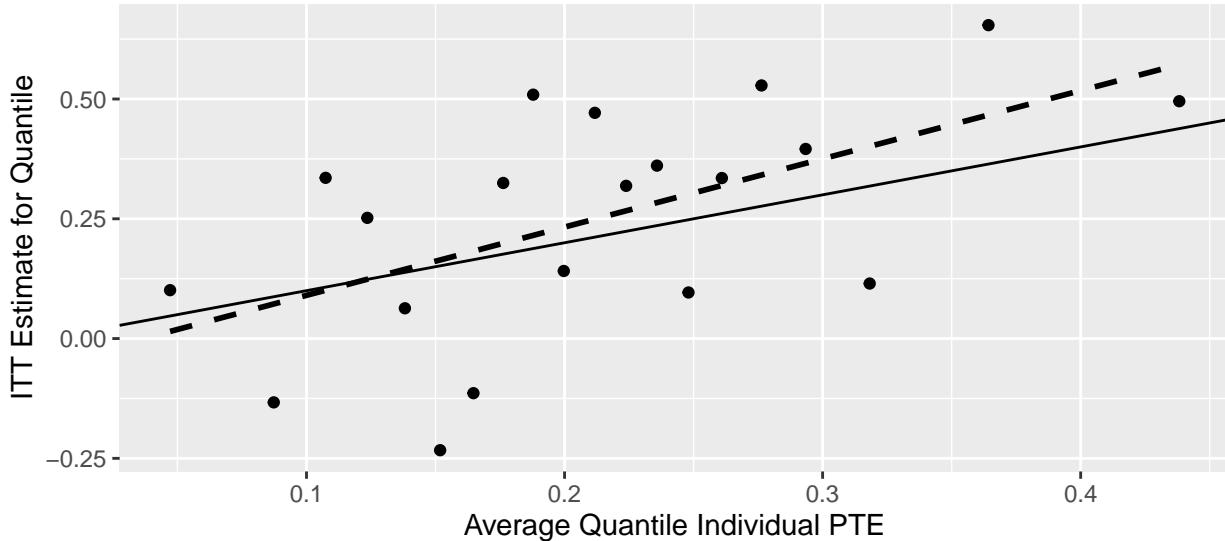
0.1722649 26 1 0.694016 1 0.02760692 0.8210774



## 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 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 quartilesOur causal forest estimates an overall average treatment effect of 0.216 (0.029).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.017 and a ‘differential forest prediction’ (DFP) coefficient of 0.578, with a corresponding DFP p-value of 0.045.<sup>2</sup>

### Calibration Plot: Math GPA



<sup>2</sup>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	0.305 +/- 0.119	0.060
Individual PTE Quartile 3	0.298 +/- 0.11	0.056
Individual PTE Quartile 2	0.137 +/- 0.115	0.059
Individual PTE Quartile 1	0.125 +/- 0.115	0.059
Top 2 PTE Quartiles	0.301 +/- 0.081	0.041
Bottom 2 PTE Quartiles	0.132 +/- 0.081	0.041
Bottom 3 PTE Quartiles	0.186 +/- 0.065	0.033

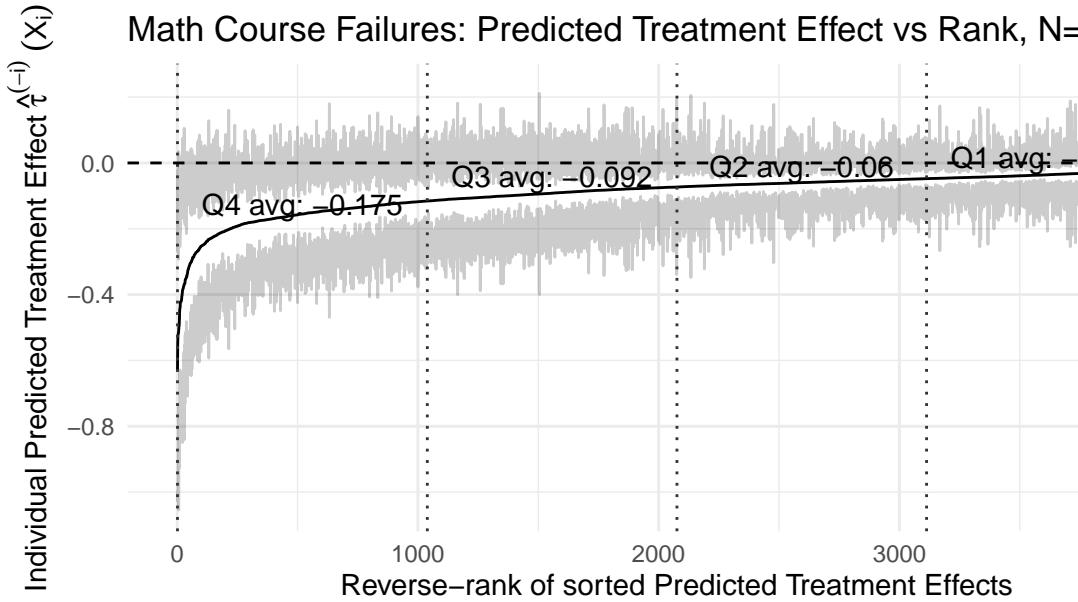
Table 5: Differences between subsample average treatment effects: Math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.017, 0.254]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[0.054, 0.283]
Quartile 4	Quartile 1	[0.014, 0.345]

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 $\hat{\tau}$	0.101	0.176	0.236	0.338
N	1038.000	1037.000	1037.000	1037.000
Age	14.300	14.499	14.798	14.993
Female	0.187	0.205	0.135	0.096
Has IEP	0.189	0.184	0.204	0.113
Has Free/Reduced Lunch	0.976	0.908	0.865	0.804
Black	0.537	0.601	0.545	0.393
Hispanic	0.430	0.361	0.421	0.534
Other Race	0.034	0.039	0.034	0.072
In 9th Grade	0.988	0.883	0.579	0.350
In 10th Grade	0.012	0.108	0.412	0.647
Baseline GPA	2.318	2.185	2.349	2.134
Num. A's	9.564	7.640	6.828	2.017
Num. B's	8.593	7.832	7.272	3.177
Num. C's	11.149	9.255	6.879	3.603
Num. D's	5.679	3.388	2.212	2.062
Num. F's	3.916	3.953	2.334	0.862
Missing Baseline GPA/Grades	0.000	0.000	0.001	0.318
Days Absent	15.046	20.274	17.813	16.435
Missing Attendance Data	0.000	0.000	0.000	0.178
Math Test Score (Z)	-0.071	-0.073	-0.066	0.073
Reading Test Score (Z)	0.022	-0.067	-0.079	0.000
Missing Math Test	0.092	0.146	0.064	0.213
Missing Reading Test	0.096	0.152	0.066	0.214
Out-of-School Suspensions	1.016	2.080	1.527	0.626
Disciplinary Incidents	0.652	1.484	1.219	0.644
Any Arrests at Baseline	0.116	0.214	0.180	0.109
Arrests: Violent Crime	0.075	0.195	0.101	0.065
Arrests: Property Crime	0.037	0.105	0.069	0.047
Arrests: Drug Crime	0.031	0.063	0.077	0.036
Math Score - Decile in Previous School	5.593	5.564	5.733	6.249
In Study 2	0.610	0.541	0.435	0.294
Participated in Year 1 of Study	0.282	0.278	0.232	0.256

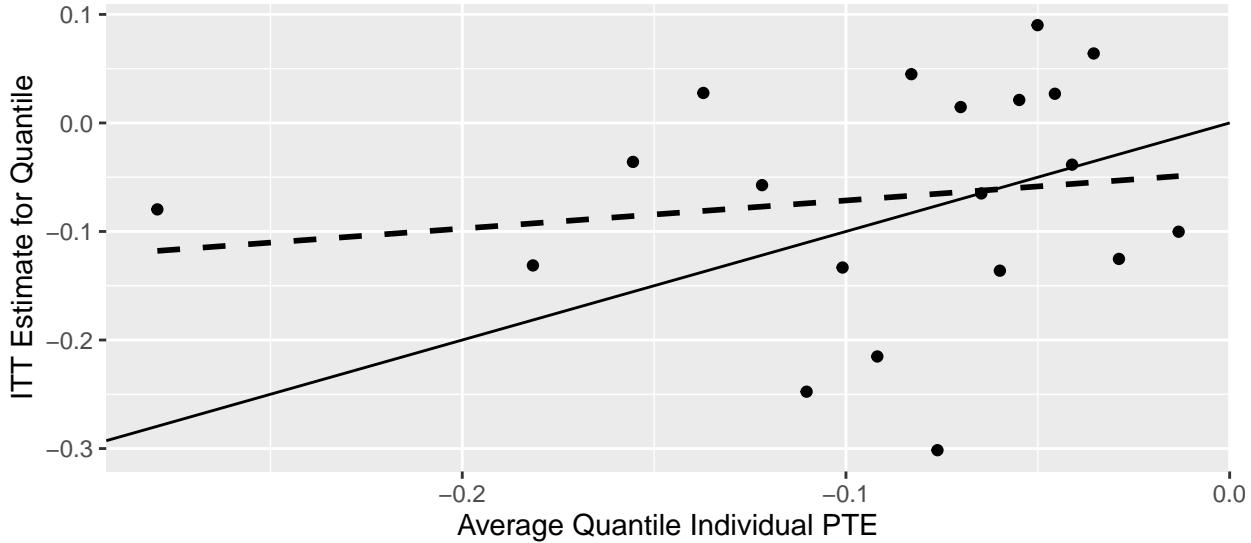
0.4242901 13 4 0.7514214 1 0.03659856 1.807611



## 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 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 quartilesOur causal forest estimates an overall average treatment effect of -0.09 (0.02).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.012 and a ‘differential forest prediction’ (DFP) coefficient of 0.848, with a corresponding DFP p-value of 0.049.<sup>3</sup>

### Calibration Plot: Math Course Failures



<sup>3</sup>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	-0.09 +/- 0.039	0.020
Individual PTE Quartile 4	-0.098 +/- 0.102	0.052
Individual PTE Quartile 3	-0.165 +/- 0.083	0.042
Individual PTE Quartile 2	-0.057 +/- 0.064	0.033
Individual PTE Quartile 1	-0.046 +/- 0.053	0.027
Top 2 PTE Quartiles	-0.133 +/- 0.066	0.034
Bottom 2 PTE Quartiles	-0.051 +/- 0.042	0.021
Bottom 3 PTE Quartiles	-0.087 +/- 0.039	0.020

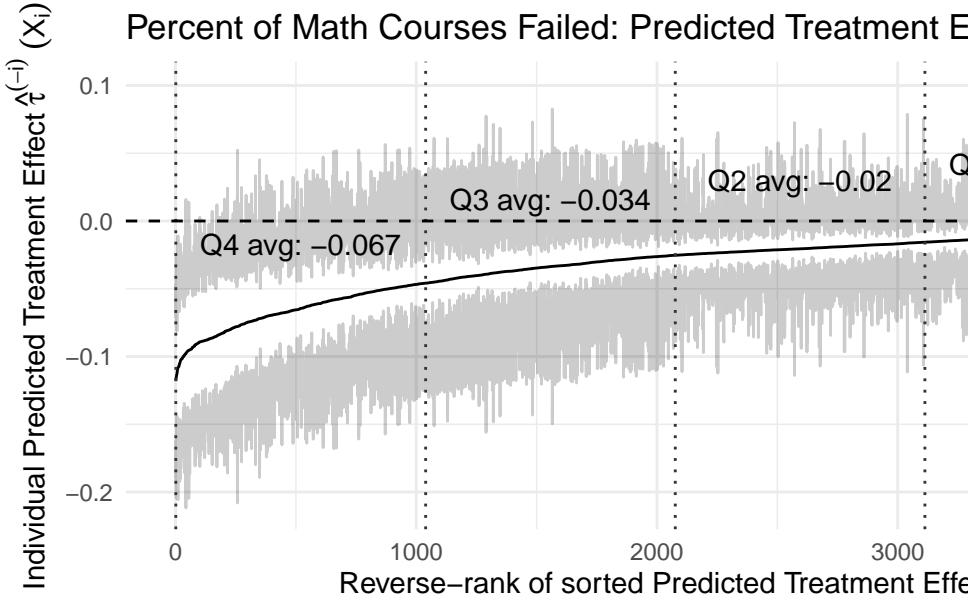
Table 8: Differences between subsample average treatment effects: Math Course Failures

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.12, 0.099]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.16, -0.004]
Quartile 4	Quartile 1	[-0.167, 0.064]

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 $\hat{\tau}$	-0.033	-0.060	-0.092	-0.175
N	1037.000	1037.000	1037.000	1038.000
Age	14.343	14.472	14.752	15.022
Female	0.182	0.187	0.155	0.099
Has IEP	0.142	0.207	0.183	0.158
Has Free/Reduced Lunch	0.987	0.916	0.768	0.882
Black	0.518	0.484	0.541	0.533
Hispanic	0.462	0.467	0.376	0.441
Other Race	0.020	0.049	0.083	0.026
In 9th Grade	0.873	0.763	0.601	0.565
In 10th Grade	0.127	0.234	0.394	0.422
Baseline GPA	2.831	2.710	2.140	1.202
Num. A's	11.497	9.062	4.208	1.291
Num. B's	10.628	9.438	4.804	2.012
Num. C's	9.878	9.706	6.735	4.574
Num. D's	2.784	2.498	3.201	4.859
Num. F's	0.970	1.223	2.336	6.534
Missing Baseline GPA/Grades	0.001	0.015	0.178	0.124
Days Absent	10.173	10.044	18.578	31.254
Missing Attendance Data	0.000	0.007	0.158	0.013
Math Test Score (Z)	0.522	-0.083	-0.247	-0.421
Reading Test Score (Z)	0.372	-0.017	-0.151	-0.396
Missing Math Test	0.068	0.038	0.237	0.171
Missing Reading Test	0.071	0.040	0.238	0.179
Out-of-School Suspensions	0.307	0.396	1.371	3.326
Disciplinary Incidents	0.203	0.290	1.101	2.499
Any Arrests at Baseline	0.104	0.065	0.154	0.296
Arrests: Violent Crime	0.052	0.028	0.124	0.231
Arrests: Property Crime	0.033	0.020	0.055	0.150
Arrests: Drug Crime	0.016	0.017	0.038	0.135
Math Score - Decile in Previous School	7.396	5.384	5.184	4.876
In Study 2	0.568	0.534	0.443	0.335
Participated in Year 1 of Study	0.243	0.270	0.278	0.257

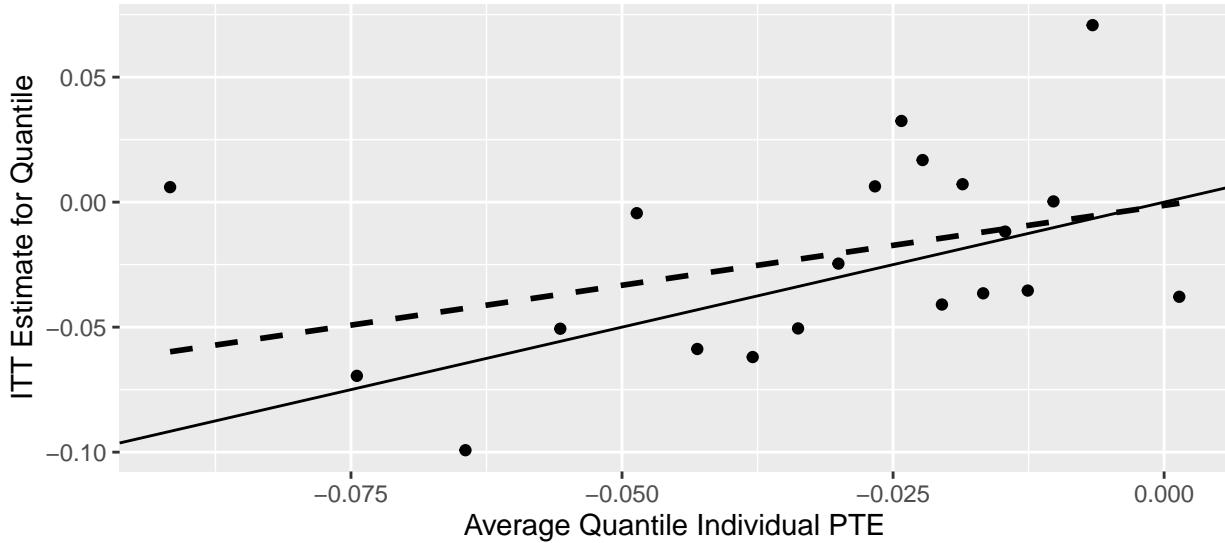
0.4610224 14 2 0.7945048 1 0.0180163 1.929979



## 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 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 quartilesOur causal forest estimates an overall average treatment effect of -0.033 (0.009).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 0.38, with a corresponding DFP p-value of 0.207.<sup>4</sup>

### Calibration Plot: Percent of Math Courses Failed



<sup>4</sup>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 +/- 0.018	0.009
Individual PTE Quartile 4	-0.054 +/- 0.045	0.023
Individual PTE Quartile 3	-0.043 +/- 0.036	0.018
Individual PTE Quartile 2	-0.019 +/- 0.025	0.013
Individual PTE Quartile 1	-0.017 +/- 0.031	0.016
Top 2 PTE Quartiles	-0.048 +/- 0.029	0.015
Bottom 2 PTE Quartiles	-0.018 +/- 0.02	0.010
Bottom 3 PTE Quartiles	-0.026 +/- 0.018	0.009

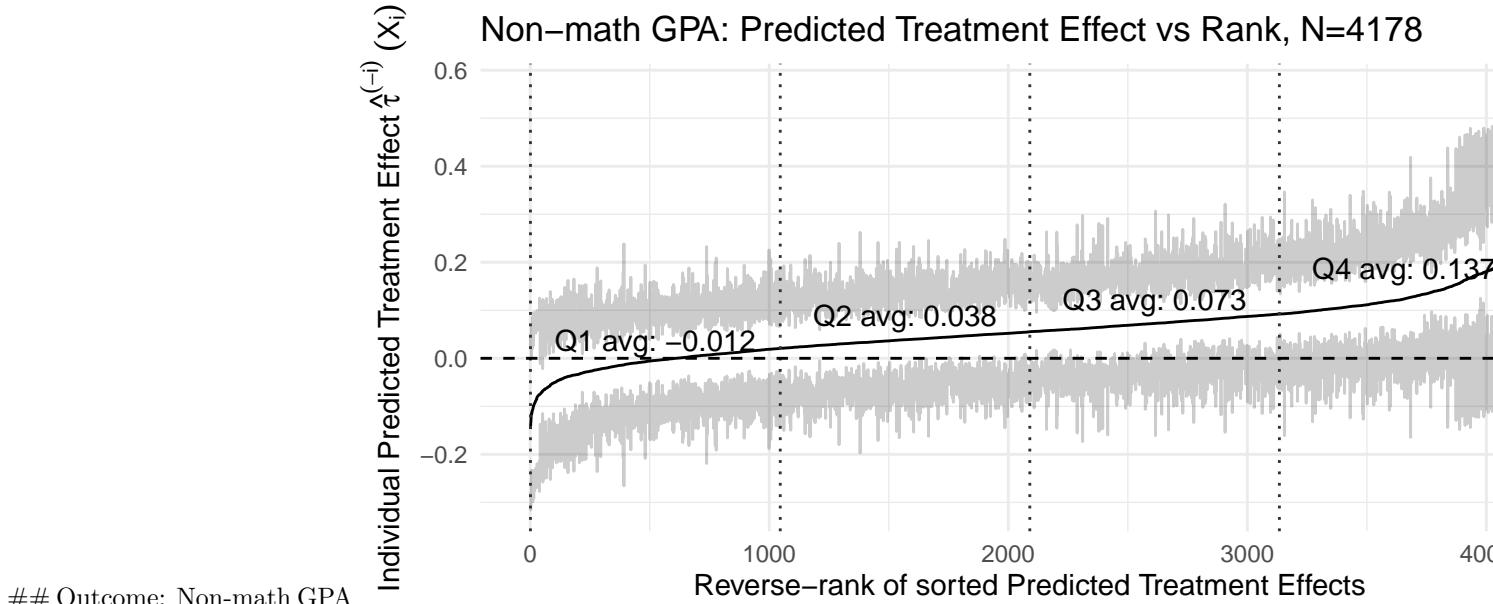
Table 11: Differences between subsample average treatment effects: Percent of Math Courses Failed

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.076, 0.021]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.065, 0.005]
Quartile 4	Quartile 1	[-0.091, 0.018]

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

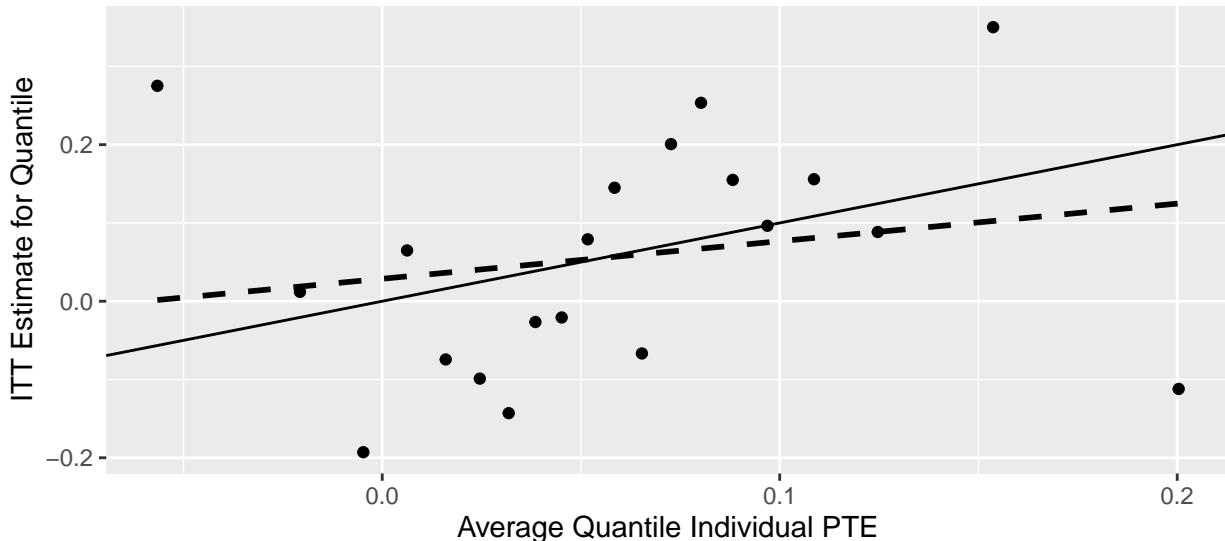
Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.009	-0.020	-0.034	-0.067
N	1037.000	1037.000	1037.000	1038.000
Age	14.287	14.393	14.696	15.212
Female	0.198	0.200	0.143	0.084
Has IEP	0.155	0.191	0.191	0.153
Has Free/Reduced Lunch	0.991	0.958	0.740	0.865
Black	0.539	0.479	0.526	0.532
Hispanic	0.448	0.478	0.398	0.421
Other Race	0.013	0.042	0.076	0.047
In 9th Grade	0.983	0.812	0.622	0.384
In 10th Grade	0.017	0.186	0.374	0.600
Baseline GPA	2.570	2.732	2.358	1.245
Num. A's	10.610	9.747	4.831	0.869
Num. B's	10.414	9.779	5.436	1.253
Num. C's	11.423	9.695	6.327	3.448
Num. D's	4.077	2.402	2.715	4.148
Num. F's	2.520	1.528	2.237	4.778
Missing Baseline GPA/Grades	0.000	0.014	0.161	0.144
Days Absent	12.278	9.836	15.139	32.577
Missing Attendance Data	0.000	0.009	0.138	0.032
Math Test Score (Z)	0.303	-0.022	-0.185	-0.309
Reading Test Score (Z)	0.165	0.029	-0.084	-0.283
Missing Math Test	0.061	0.034	0.236	0.184
Missing Reading Test	0.065	0.037	0.239	0.188
Out-of-School Suspensions	0.365	0.433	1.333	3.297
Disciplinary Incidents	0.240	0.301	0.936	2.621
Any Arrests at Baseline	0.149	0.057	0.114	0.299
Arrests: Violent Crime	0.086	0.044	0.093	0.213
Arrests: Property Crime	0.043	0.023	0.060	0.132
Arrests: Drug Crime	0.027	0.010	0.038	0.132
Math Score - Decile in Previous School	6.767	5.598	5.288	5.241
In Study 2	0.635	0.526	0.460	0.260
Participated in Year 1 of Study	0.271	0.262	0.271	0.244

0.3079006 19 3 0.640886 1 0.007615928 0.2416616



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 quartilesOur causal forest estimates an overall average treatment effect of 0.058 (0.023).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1 and a ‘differential forest prediction’ (DFP) coefficient of -0.074, with a corresponding DFP p-value of 0.566.<sup>5</sup>

### Calibration Plot: Non-math GPA



<sup>5</sup>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	0.041 +/- 0.097	0.050
Individual PTE Quartile 3	0.167 +/- 0.086	0.044
Individual PTE Quartile 2	0.009 +/- 0.088	0.045
Individual PTE Quartile 1	0.014 +/- 0.09	0.046
Top 2 PTE Quartiles	0.105 +/- 0.065	0.033
Bottom 2 PTE Quartiles	0.012 +/- 0.063	0.032
Bottom 3 PTE Quartiles	0.064 +/- 0.051	0.026

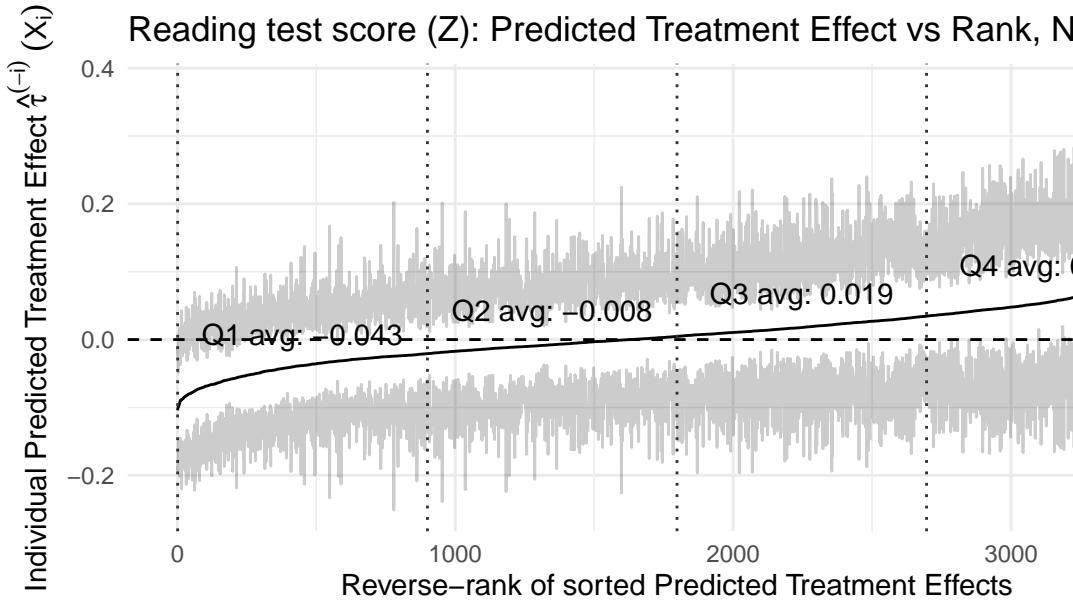
Table 14: Differences between subsample average treatment effects: Non-math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.132, 0.088]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[0.002, 0.184]
Quartile 4	Quartile 1	[-0.105, 0.161]

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 $\hat{\tau}$	-0.012	0.038	0.073	0.137
N	1045.000	1044.000	1044.000	1045.000
Age	14.611	14.559	14.579	14.862
Female	0.095	0.177	0.205	0.146
Has IEP	0.155	0.175	0.202	0.157
Has Free/Reduced Lunch	0.926	0.903	0.867	0.859
Black	0.404	0.486	0.568	0.622
Hispanic	0.553	0.467	0.386	0.337
Other Race	0.043	0.047	0.046	0.041
In 9th Grade	0.686	0.734	0.764	0.612
In 10th Grade	0.313	0.264	0.228	0.376
Baseline GPA	2.350	2.346	2.417	1.757
Num. A's	6.193	7.420	8.626	3.692
Num. B's	8.304	7.940	7.488	3.016
Num. C's	10.118	9.456	7.483	3.730
Num. D's	4.099	3.784	2.738	2.744
Num. F's	1.833	2.507	2.924	3.944
Missing Baseline GPA/Grades	0.001	0.008	0.043	0.266
Days Absent	12.752	13.424	15.661	31.417
Missing Attendance Data	0.000	0.000	0.000	0.178
Math Test Score (Z)	0.152	-0.080	-0.086	-0.138
Reading Test Score (Z)	0.249	-0.091	-0.136	-0.142
Missing Math Test	0.167	0.025	0.060	0.271
Missing Reading Test	0.171	0.028	0.061	0.277
Out-of-School Suspensions	0.259	0.425	0.996	4.368
Disciplinary Incidents	0.277	0.406	0.794	3.017
Any Arrests at Baseline	0.062	0.105	0.144	0.319
Arrests: Violent Crime	0.025	0.057	0.114	0.254
Arrests: Property Crime	0.017	0.034	0.060	0.150
Arrests: Drug Crime	0.014	0.012	0.037	0.146
Math Score - Decile in Previous School	6.286	5.573	5.639	5.591
In Study 2	0.447	0.515	0.526	0.390
Participated in Year 1 of Study	0.287	0.289	0.240	0.226

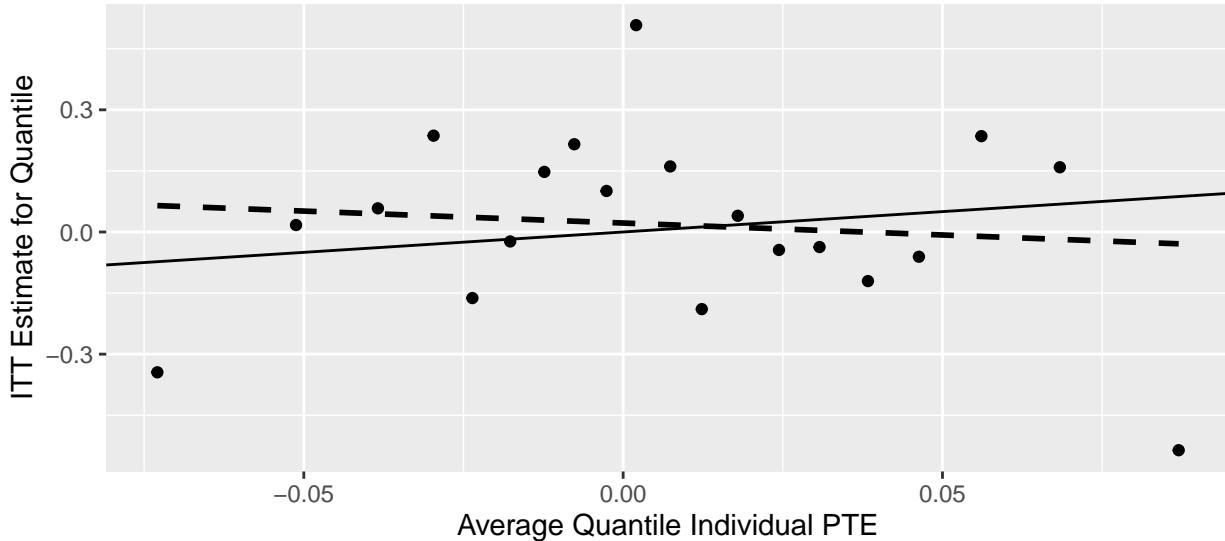
0.4436948 15 3 0.5595897 1 0.02643746 0.8033093



## Outcome: Reading 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 quartilesOur causal forest estimates an overall average treatment effect of 0.008 (0.027).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.131 and a ‘differential forest prediction’ (DFP) coefficient of -1.648, with a corresponding DFP p-value of 0.983.<sup>6</sup>

### Calibration Plot: Reading test score (Z)



<sup>6</sup>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	0.008 +/- 0.053	0.027
Individual PTE Quartile 4	-0.089 +/- 0.126	0.064
Individual PTE Quartile 3	0.03 +/- 0.108	0.055
Individual PTE Quartile 2	0.087 +/- 0.098	0.050
Individual PTE Quartile 1	0.002 +/- 0.093	0.048
Top 2 PTE Quartiles	-0.03 +/- 0.083	0.042
Bottom 2 PTE Quartiles	0.046 +/- 0.067	0.034
Bottom 3 PTE Quartiles	0.04 +/- 0.058	0.029

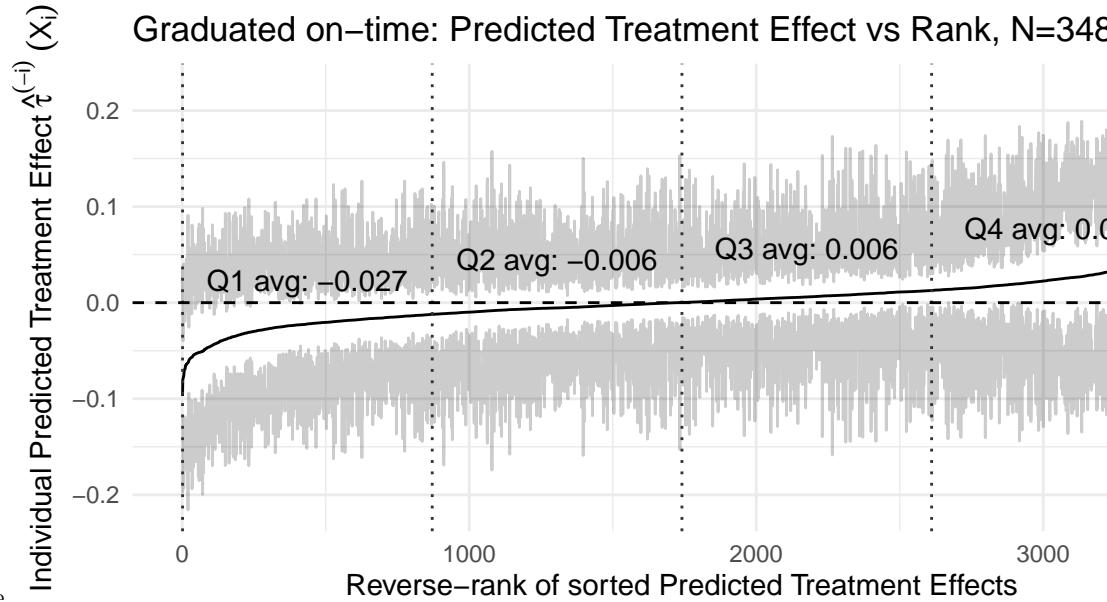
Table 17: Differences between subsample average treatment effects: Reading test score (Z)

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.268, 0.008]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.182, 0.031]
Quartile 4	Quartile 1	[-0.248, 0.065]

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 $\hat{\tau}$	-0.043	-0.008	0.019	0.059
N	899.000	898.000	898.000	899.000
Age	14.662	14.639	14.575	14.313
Female	0.116	0.148	0.178	0.196
Has IEP	0.390	0.177	0.072	0.038
Has Free/Reduced Lunch	0.953	0.902	0.865	0.849
Black	0.493	0.546	0.539	0.464
Hispanic	0.497	0.418	0.419	0.449
Other Race	0.010	0.037	0.042	0.087
In 9th Grade	0.671	0.668	0.704	0.871
In 10th Grade	0.325	0.332	0.295	0.128
Baseline GPA	2.032	2.251	2.582	2.749
Num. A's	4.623	6.076	7.886	10.833
Num. B's	5.736	6.844	8.639	8.643
Num. C's	9.181	8.830	8.269	6.677
Num. D's	5.051	3.584	2.480	1.957
Num. F's	3.592	2.507	1.772	1.536
Missing Baseline GPA/Grades	0.012	0.042	0.061	0.160
Days Absent	17.075	15.409	11.660	10.073
Missing Attendance Data	0.000	0.000	0.000	0.135
Math Test Score (Z)	-0.809	-0.225	0.432	0.928
Reading Test Score (Z)	-0.670	-0.162	0.422	0.721
Missing Math Test	0.000	0.008	0.041	0.280
Missing Reading Test	0.000	0.011	0.047	0.283
Out-of-School Suspensions	0.959	1.052	0.899	0.609
Disciplinary Incidents	0.954	0.851	0.566	0.440
Any Arrests at Baseline	0.118	0.127	0.095	0.073
Arrests: Violent Crime	0.063	0.073	0.062	0.038
Arrests: Property Crime	0.036	0.036	0.024	0.023
Arrests: Drug Crime	0.021	0.033	0.012	0.012
Math Score - Decile in Previous School	3.539	5.204	7.060	8.427
In Study 2	0.434	0.433	0.513	0.576
Participated in Year 1 of Study	0.297	0.275	0.242	0.254

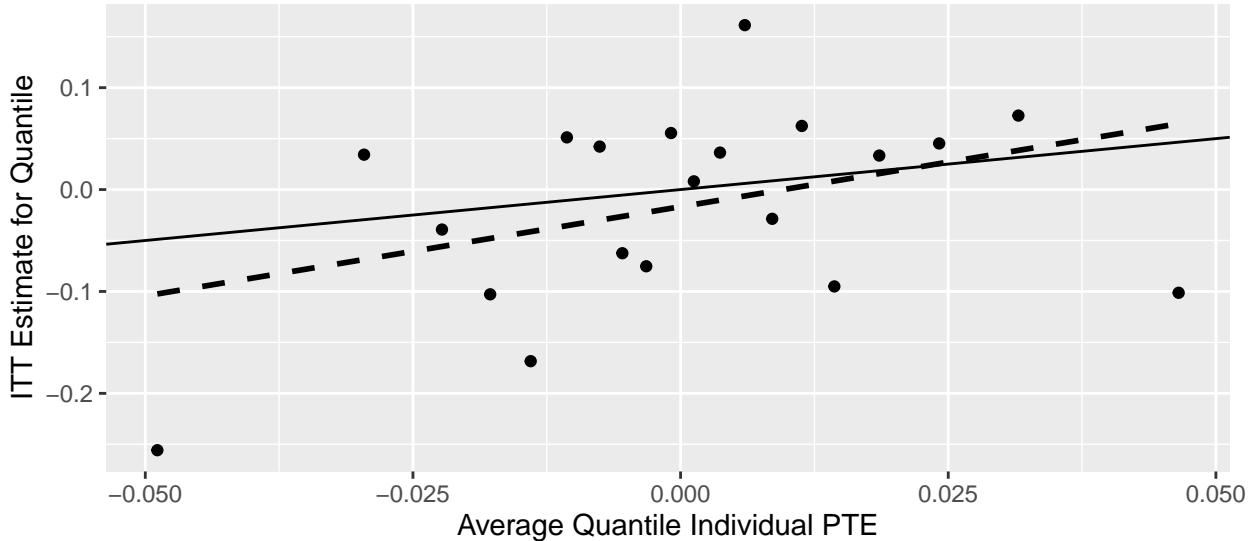
0.2472334 24 2 0.5302957 0 0.009221531 2.414188



## 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 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 quartilesOur causal forest estimates an overall average treatment effect of 0.001 (0.012).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 5.962 and a ‘differential forest prediction’ (DFP) coefficient of -0.683, with a corresponding DFP p-value of 0.844.<sup>7</sup>

### Calibration Plot: Graduated on-time



<sup>7</sup>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	-0.03 +/- 0.052	0.026
Individual PTE Quartile 3	0.04 +/- 0.046	0.023
Individual PTE Quartile 2	0.008 +/- 0.045	0.023
Individual PTE Quartile 1	-0.014 +/- 0.052	0.027
Top 2 PTE Quartiles	0.005 +/- 0.035	0.018
Bottom 2 PTE Quartiles	-0.002 +/- 0.035	0.018
Bottom 3 PTE Quartiles	0.012 +/- 0.028	0.014

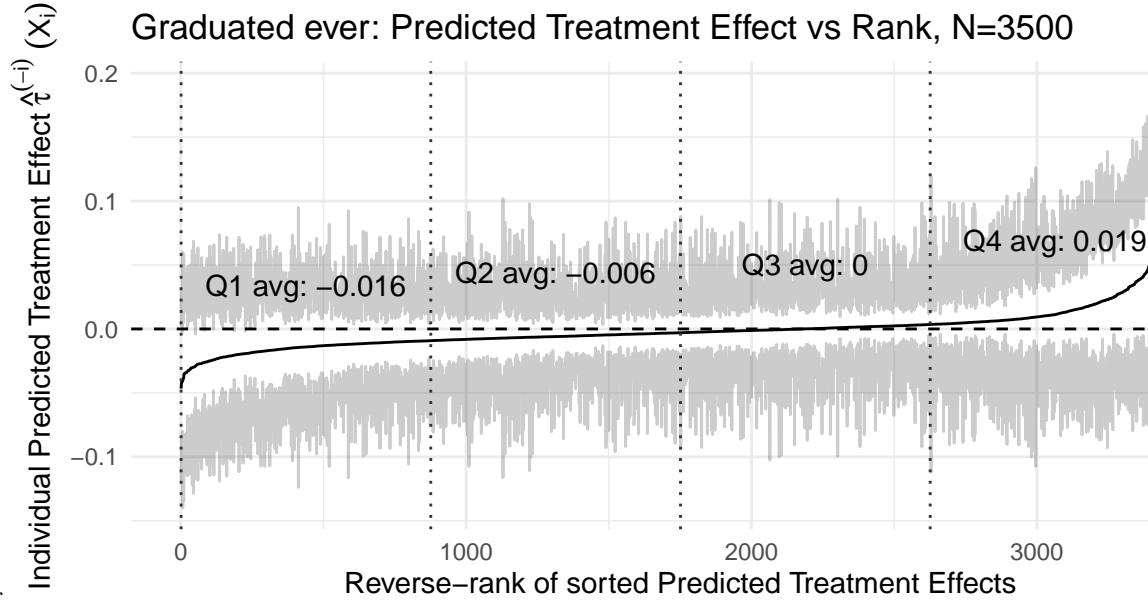
Table 20: Differences between subsample average treatment effects: Graduated on-time

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.1, 0.017]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.041, 0.057]
Quartile 4	Quartile 1	[-0.089, 0.058]

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 $\hat{\tau}$	-0.027	-0.006	0.006	0.027
N	870.000	870.000	870.000	870.000
Age	14.910	14.560	14.338	14.537
Female	0.092	0.171	0.184	0.179
Has IEP	0.187	0.206	0.140	0.164
Has Free/Reduced Lunch	0.864	0.900	0.890	0.898
Black	0.514	0.547	0.478	0.622
Hispanic	0.431	0.425	0.477	0.346
Other Race	0.055	0.028	0.045	0.032
In 9th Grade	0.460	0.717	0.853	0.791
In 10th Grade	0.531	0.283	0.144	0.206
Baseline GPA	2.168	2.547	2.663	2.050
Num. A's	3.460	7.739	10.362	7.163
Num. B's	5.495	8.175	9.454	5.922
Num. C's	7.483	8.683	8.323	7.070
Num. D's	3.639	2.684	2.549	3.800
Num. F's	1.616	1.793	1.960	4.059
Missing Baseline GPA/Grades	0.017	0.038	0.079	0.138
Days Absent	14.998	11.758	11.028	24.023
Missing Attendance Data	0.000	0.003	0.025	0.095
Math Test Score (Z)	-0.130	-0.093	0.256	0.051
Reading Test Score (Z)	-0.068	-0.059	0.196	-0.029
Missing Math Test	0.133	0.057	0.059	0.163
Missing Reading Test	0.138	0.063	0.059	0.170
Out-of-School Suspensions	0.093	0.208	0.343	4.241
Disciplinary Incidents	0.167	0.298	0.351	2.910
Any Arrests at Baseline	0.100	0.085	0.076	0.277
Arrests: Violent Crime	0.036	0.054	0.059	0.252
Arrests: Property Crime	0.032	0.055	0.038	0.143
Arrests: Drug Crime	0.021	0.026	0.024	0.094
Math Score - Decile in Previous School	5.442	5.532	6.550	6.128
In Study 2	0.364	0.489	0.547	0.523
Participated in Year 1 of Study	0.286	0.228	0.236	0.244

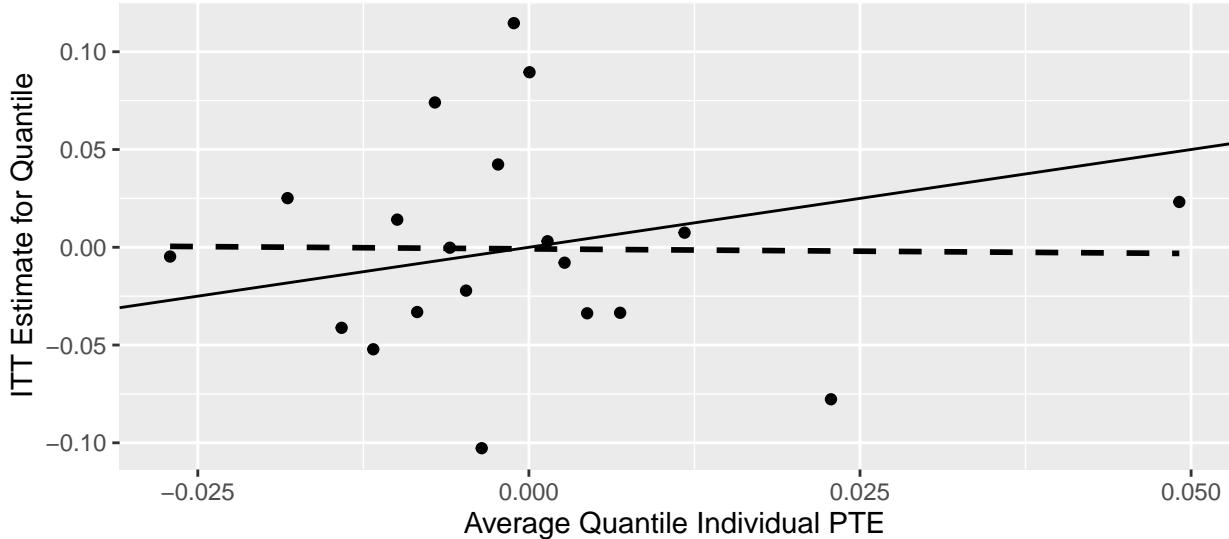
0.1615604 41 1 0.7536215 1 0.008674368 0.1192588



## Outcome: Graduated ever

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 (0.011). We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 0.858 and a ‘differential forest prediction’ (DFP) coefficient of -1.301, with a corresponding DFP p-value of 0.913.<sup>8</sup>

### Calibration Plot: Graduated ever



<sup>8</sup>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 +/- 0.022	0.011
Individual PTE Quartile 4	-0.027 +/- 0.052	0.026
Individual PTE Quartile 3	0.008 +/- 0.037	0.019
Individual PTE Quartile 2	-0.008 +/- 0.039	0.020
Individual PTE Quartile 1	0.025 +/- 0.048	0.024
Top 2 PTE Quartiles	-0.009 +/- 0.032	0.016
Bottom 2 PTE Quartiles	0.008 +/- 0.031	0.016
Bottom 3 PTE Quartiles	0.008 +/- 0.024	0.012

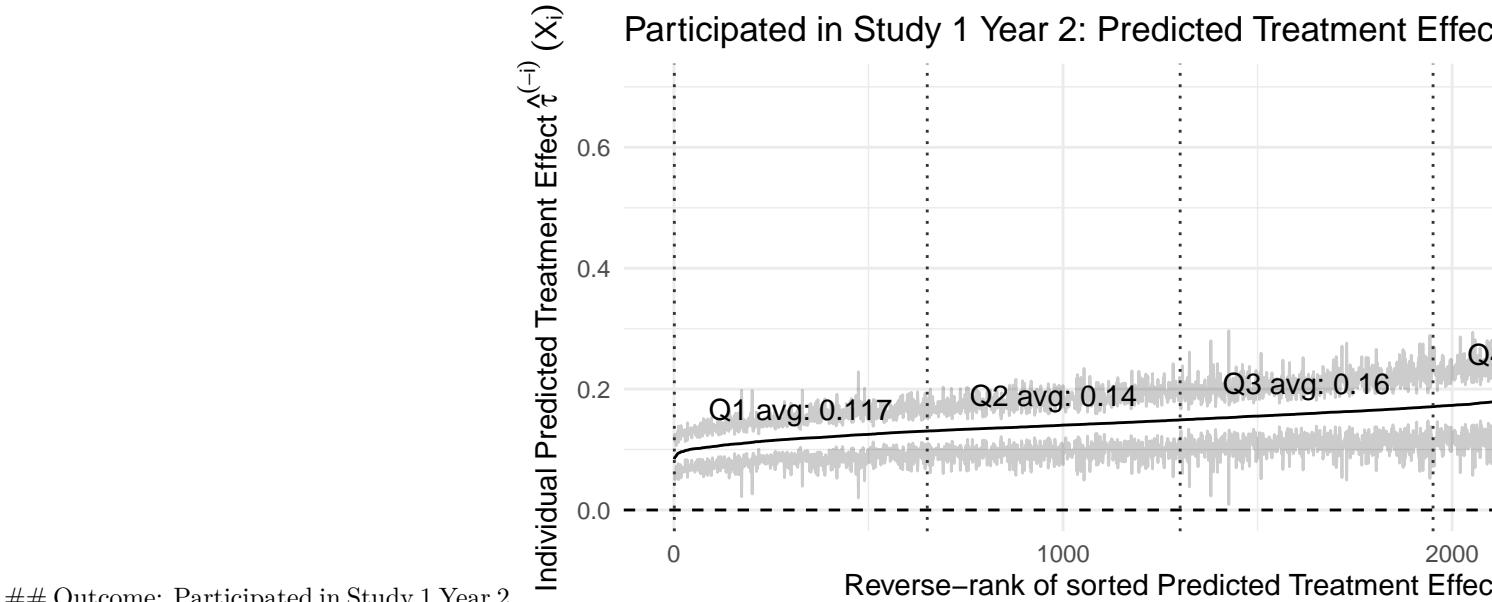
Table 23: Differences between subsample average treatment effects: Graduated ever

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.092, 0.022]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.062, 0.027]
Quartile 4	Quartile 1	[-0.121, 0.019]

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 $\hat{\tau}$	-0.016	-0.006	0.000	0.019
N	875.000	875.000	875.000	875.000
Age	14.599	14.715	14.333	14.689
Female	0.134	0.149	0.190	0.153
Has IEP	0.155	0.179	0.157	0.206
Has Free/Reduced Lunch	0.873	0.883	0.891	0.902
Black	0.597	0.495	0.489	0.581
Hispanic	0.363	0.465	0.465	0.384
Other Race	0.040	0.040	0.046	0.035
In 9th Grade	0.753	0.542	0.811	0.699
In 10th Grade	0.238	0.451	0.182	0.286
Baseline GPA	2.015	2.580	2.799	2.015
Num. A's	4.551	6.613	11.553	5.998
Num. B's	6.059	7.184	9.789	5.957
Num. C's	9.918	7.606	7.431	6.606
Num. D's	5.179	2.629	1.960	2.944
Num. F's	3.697	1.298	1.165	3.353
Missing Baseline GPA/Grades	0.002	0.019	0.074	0.177
Days Absent	14.050	10.460	10.642	27.538
Missing Attendance Data	0.000	0.000	0.000	0.123
Math Test Score (Z)	0.117	0.077	0.216	-0.487
Reading Test Score (Z)	0.010	0.067	0.238	-0.398
Missing Math Test	0.027	0.027	0.031	0.327
Missing Reading Test	0.032	0.026	0.035	0.336
Out-of-School Suspensions	0.278	0.280	0.475	3.947
Disciplinary Incidents	0.200	0.281	0.424	2.894
Any Arrests at Baseline	0.142	0.061	0.062	0.278
Arrests: Violent Crime	0.069	0.039	0.033	0.269
Arrests: Property Crime	0.039	0.040	0.032	0.157
Arrests: Drug Crime	0.021	0.024	0.022	0.109
Math Score - Decile in Previous School	6.239	6.067	6.338	4.539
In Study 2	0.504	0.430	0.534	0.461
Participated in Year 1 of Study	0.262	0.243	0.242	0.240

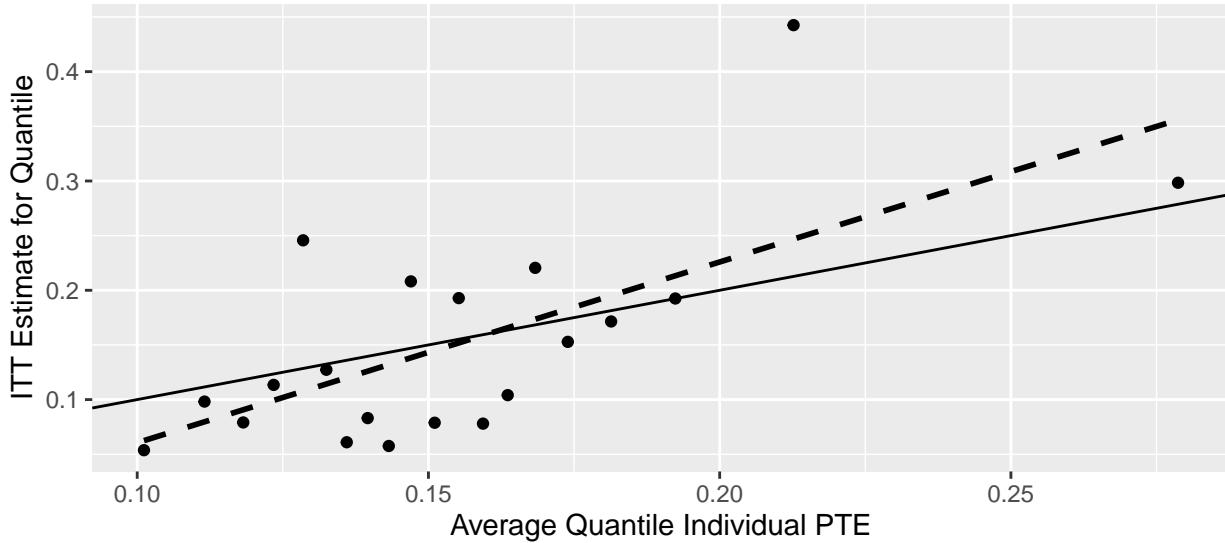
0.1612027 43 2 0.7813505 1 0.05620829 0.7390069



## 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 quartilesOur causal forest estimates an overall average treatment effect of 0.156 (0.01).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 0.999 and a ‘differential forest prediction’ (DFP) coefficient of 1.38, with a corresponding DFP p-value of 0.<sup>9</sup>

### Calibration Plot: Participated in Study 1 Year 2



<sup>9</sup>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	0.222 +/- 0.047	0.024
Individual PTE Quartile 3	0.169 +/- 0.042	0.021
Individual PTE Quartile 2	0.111 +/- 0.036	0.019
Individual PTE Quartile 1	0.12 +/- 0.036	0.018
Top 2 PTE Quartiles	0.196 +/- 0.031	0.016
Bottom 2 PTE Quartiles	0.115 +/- 0.026	0.013
Bottom 3 PTE Quartiles	0.134 +/- 0.022	0.011

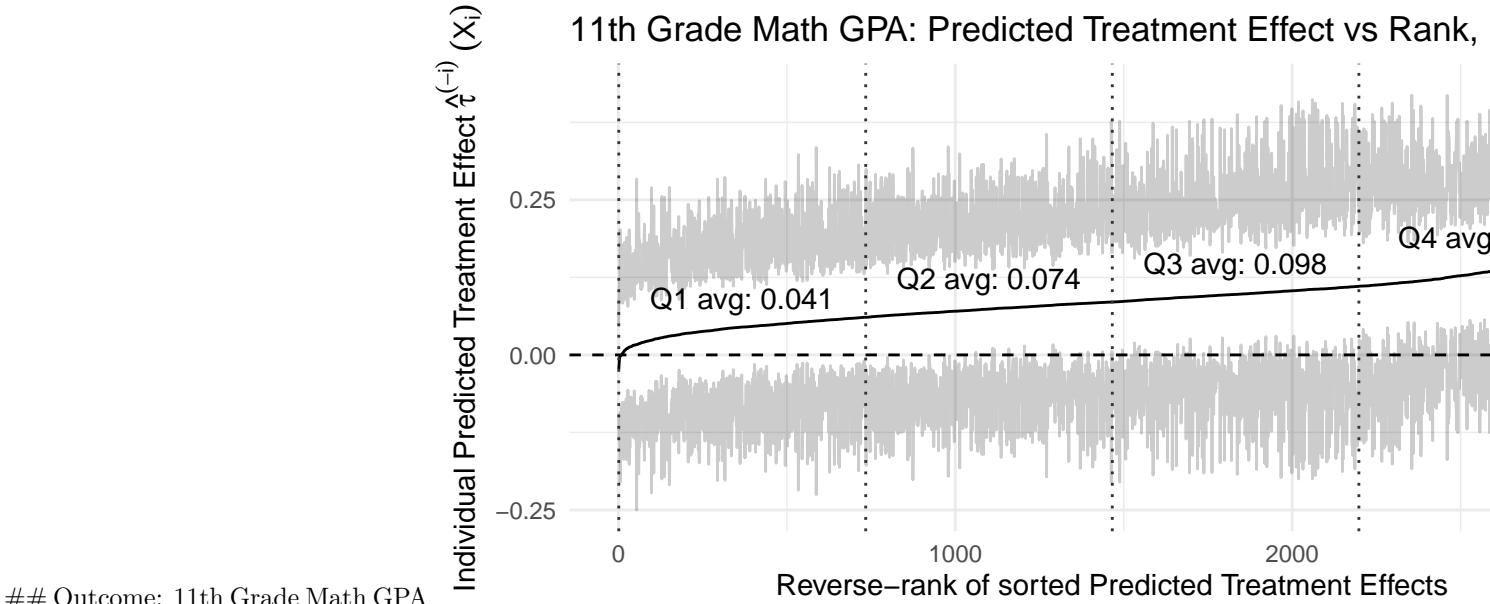
Table 26: Differences between subsample average treatment effects: Participated in Study 1 Year 2

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[0.036, 0.139]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[0.04, 0.12]
Quartile 4	Quartile 1	[0.043, 0.16]

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

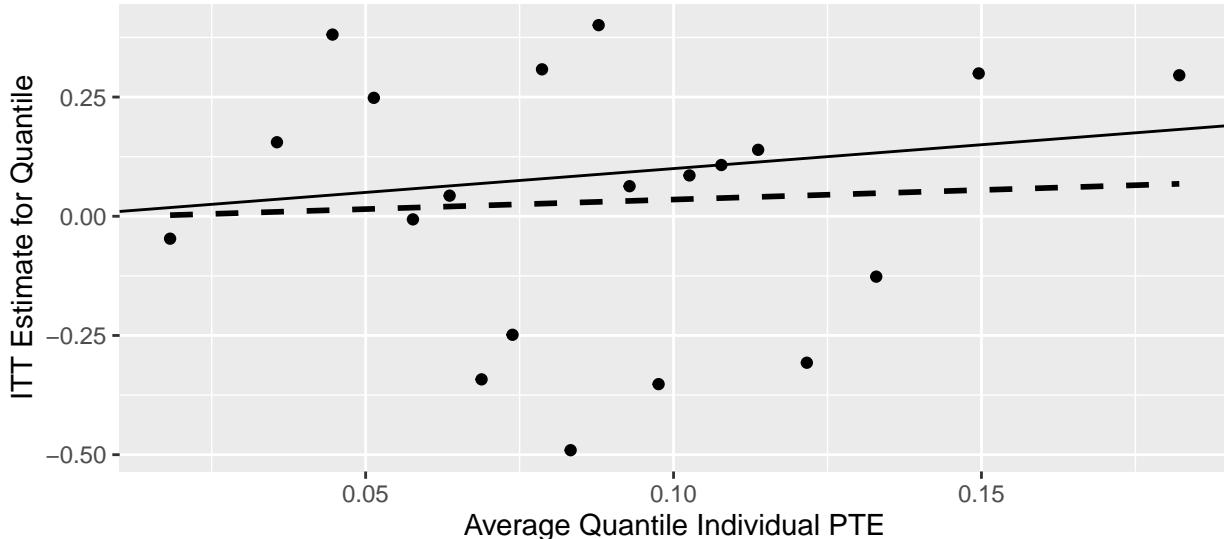
Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.117	0.140	0.160	0.208
N	650.000	650.000	650.000	650.000
Age	14.934	14.818	14.740	14.749
Female	0.002	0.000	0.002	0.002
Has IEP	0.058	0.065	0.209	0.352
Has Free/Reduced Lunch	0.860	0.894	0.882	0.826
Black	0.402	0.448	0.535	0.477
Hispanic	0.589	0.542	0.448	0.352
Other Race	0.009	0.011	0.017	0.171
In 9th Grade	0.583	0.500	0.555	0.606
In 10th Grade	0.388	0.491	0.442	0.388
Baseline GPA	2.057	2.055	2.131	2.158
Num. A's	7.309	4.406	4.822	4.360
Num. B's	5.642	5.535	5.858	5.795
Num. C's	2.485	6.091	8.494	8.908
Num. D's	1.538	3.483	4.045	4.172
Num. F's	3.671	3.125	2.769	1.991
Missing Baseline GPA/Grades	0.108	0.040	0.020	0.123
Days Absent	26.898	19.103	18.815	18.197
Missing Attendance Data	0.000	0.000	0.006	0.118
Math Test Score (Z)	0.710	0.269	-0.189	-0.874
Reading Test Score (Z)	0.649	0.130	-0.279	-0.452
Missing Math Test	0.218	0.111	0.072	0.220
Missing Reading Test	0.222	0.112	0.069	0.225
Out-of-School Suspensions	2.055	1.426	1.683	1.417
Disciplinary Incidents	1.532	1.208	1.260	1.145
Any Arrests at Baseline	0.272	0.165	0.162	0.135
Arrests: Violent Crime	0.231	0.126	0.092	0.092
Arrests: Property Crime	0.149	0.069	0.071	0.048
Arrests: Drug Crime	0.138	0.055	0.042	0.038
Math Score - Decile in Previous School	8.287	7.206	5.624	3.354
In Study 2	0.000	0.000	0.000	0.000
Participated in Year 1 of Study	0.129	0.191	0.248	0.245

0.4946198 33 1 0.664616 0 0.005445751 0.002942552



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 quartilesOur causal forest estimates an overall average treatment effect of 0.096 (0.038).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.087 and a ‘differential forest prediction’ (DFP) coefficient of -1.679, with a corresponding DFP p-value of 0.957.<sup>10</sup>

### Calibration Plot: 11th Grade Math GPA



<sup>10</sup>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	0.096 +/- 0.074	0.038
Individual PTE Quartile 4	0.005 +/- 0.15	0.076
Individual PTE Quartile 3	0.078 +/- 0.154	0.079
Individual PTE Quartile 2	0.002 +/- 0.15	0.076
Individual PTE Quartile 1	0.301 +/- 0.142	0.072
Top 2 PTE Quartiles	0.041 +/- 0.107	0.055
Bottom 2 PTE Quartiles	0.152 +/- 0.103	0.053
Bottom 3 PTE Quartiles	0.126 +/- 0.086	0.044

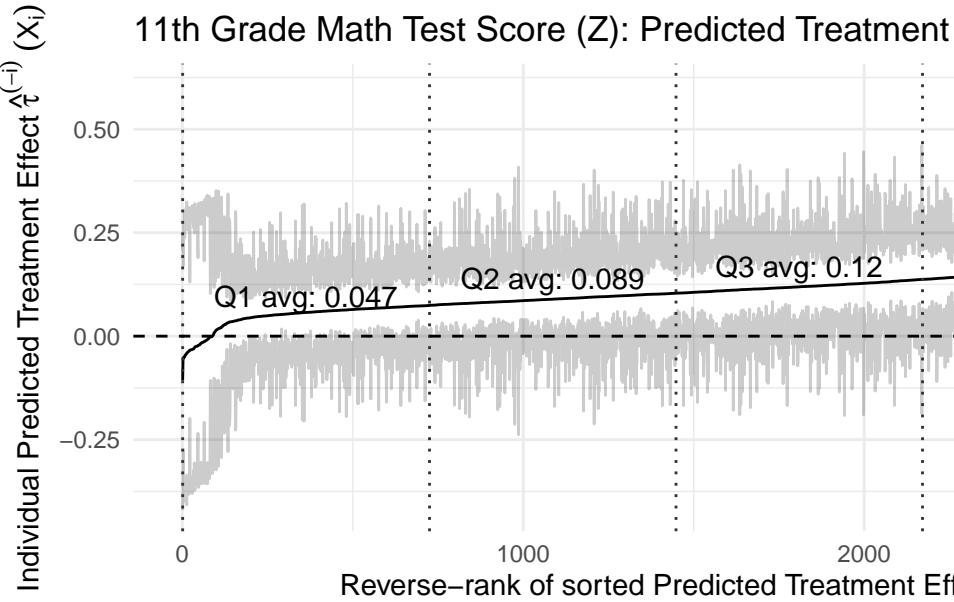
Table 29: Differences between subsample average treatment effects: 11th Grade Math GPA

Group 1	Group 2	95% Confidence Interval of Difference
Quartile 4	Bottom 3 Quartiles	[-0.294, 0.051]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.259, 0.038]
Quartile 4	Quartile 1	[-0.502, -0.09]

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 $\hat{\tau}$	0.041	0.074	0.098	0.140
N	733.000	732.000	732.000	732.000
Age	14.276	14.423	14.571	14.947
Female	0.183	0.189	0.150	0.083
Has IEP	0.095	0.105	0.209	0.306
Has Free/Reduced Lunch	0.870	0.872	0.887	0.913
Black	0.767	0.531	0.376	0.305
Hispanic	0.196	0.411	0.563	0.658
Other Race	0.037	0.057	0.061	0.037
In 9th Grade	0.910	0.810	0.708	0.313
In 10th Grade	0.087	0.187	0.288	0.682
Baseline GPA	2.476	2.624	2.527	2.297
Num. A's	8.662	9.862	8.037	4.015
Num. B's	10.117	8.641	7.216	4.701
Num. C's	10.808	7.728	7.153	6.654
Num. D's	3.364	2.620	2.807	3.367
Num. F's	2.280	2.030	1.937	1.355
Missing Baseline GPA/Grades	0.022	0.068	0.123	0.042
Days Absent	13.814	12.449	11.566	11.742
Missing Attendance Data	0.000	0.007	0.090	0.034
Math Test Score (Z)	0.496	0.229	-0.035	-0.465
Reading Test Score (Z)	0.560	0.207	-0.107	-0.482
Missing Math Test	0.095	0.111	0.108	0.034
Missing Reading Test	0.097	0.116	0.111	0.037
Out-of-School Suspensions	1.034	0.626	0.545	0.670
Disciplinary Incidents	0.573	0.608	0.629	0.789
Any Arrests at Baseline	0.119	0.078	0.077	0.092
Arrests: Violent Crime	0.071	0.056	0.048	0.040
Arrests: Property Crime	0.040	0.034	0.022	0.031
Arrests: Drug Crime	0.026	0.011	0.020	0.023
Math Score - Decile in Previous School	7.268	6.455	5.814	4.501
In Study 2	0.593	0.552	0.452	0.288
Participated in Year 1 of Study	0.252	0.258	0.286	0.309

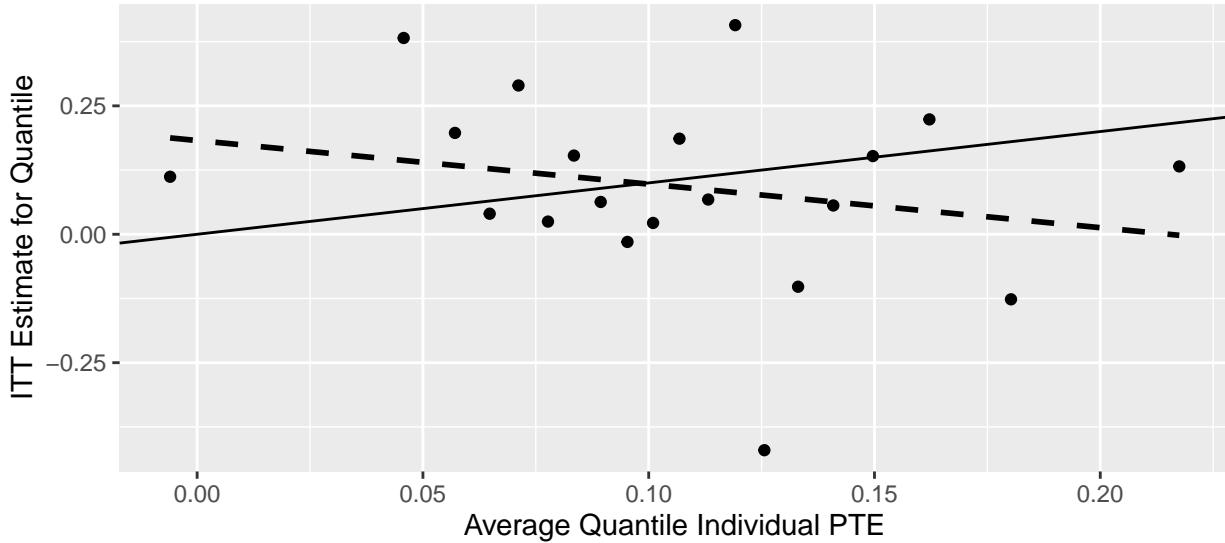
0.0556809 49 2 0.5757383 0 0.06313904 0.3481267



## 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 quartilesOur causal forest estimates an overall average treatment effect of 0.107 (0.028).We test the calibration of the forest, and estimate a ‘mean forest prediction’ (MFP) coefficient of 1.015 and a ‘differential forest prediction’ (DFP) coefficient of -1.435, with a corresponding DFP p-value of 0.982.<sup>11</sup>

### Calibration Plot: 11th Grade Math Test Score (Z)



<sup>11</sup>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	0.107 +/- 0.056	0.028
Individual PTE Quartile 4	0.035 +/- 0.116	0.059
Individual PTE Quartile 3	0.121 +/- 0.109	0.056
Individual PTE Quartile 2	0.04 +/- 0.108	0.055
Individual PTE Quartile 1	0.239 +/- 0.113	0.057
Top 2 PTE Quartiles	0.078 +/- 0.08	0.041
Bottom 2 PTE Quartiles	0.138 +/- 0.078	0.040
Bottom 3 PTE Quartiles	0.132 +/- 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	[-0.229, 0.035]
Top 2 Quartiles (3 & 4)	Bottom 2 Quartiles (1 & 2)	[-0.171, 0.051]
Quartile 4	Quartile 1	[-0.366, -0.043]

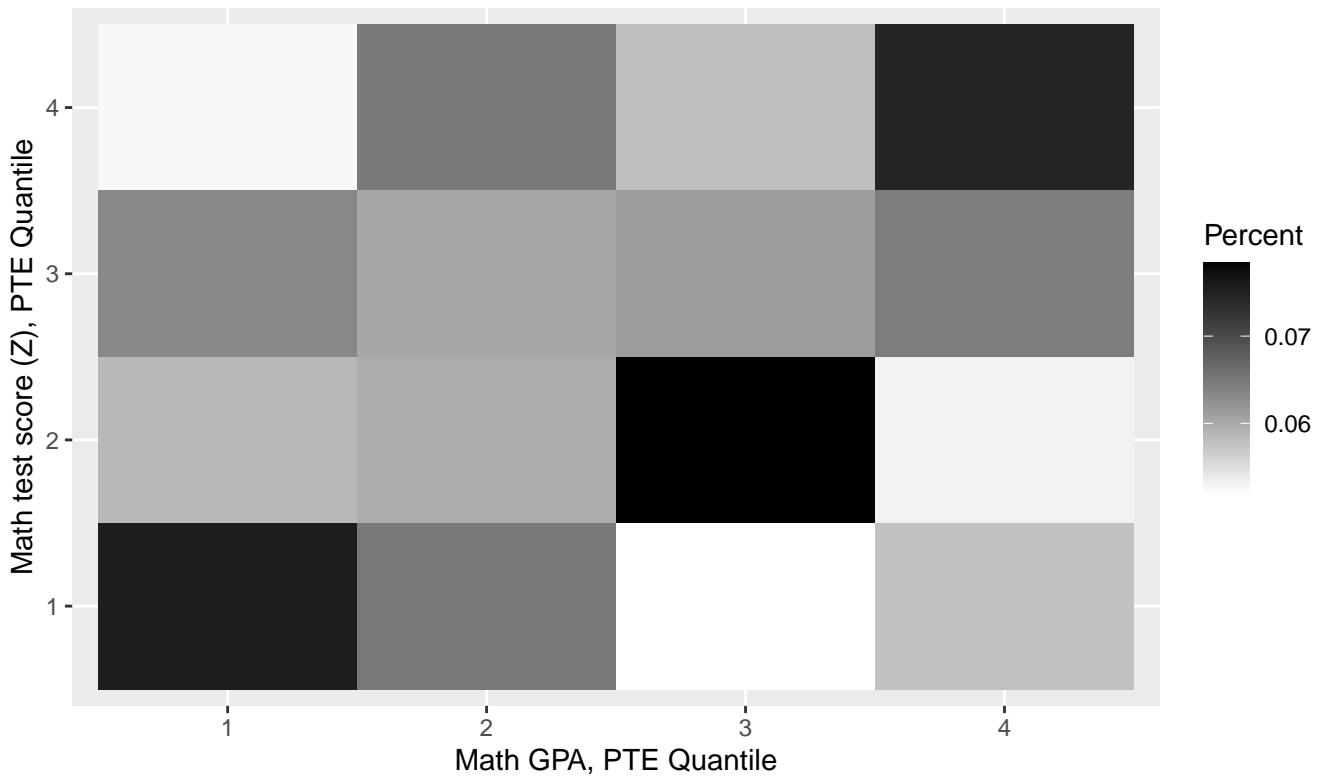
Table 33: Summary table by Quartile of Predicted Treatment Effects on 11th Grade Math Test Score (Z)

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.047	0.089	0.120	0.170
N	724.000	723.000	723.000	724.000
Age	14.526	14.505	14.562	14.500
Female	0.180	0.172	0.142	0.095
Has IEP	0.305	0.220	0.112	0.062
Has Free/Reduced Lunch	0.896	0.906	0.855	0.884
Black	0.570	0.517	0.499	0.395
Hispanic	0.396	0.441	0.445	0.554
Other Race	0.033	0.041	0.055	0.051
In 9th Grade	0.805	0.772	0.617	0.593
In 10th Grade	0.195	0.228	0.383	0.407
Baseline GPA	2.236	2.444	2.681	2.646
Num. A's	5.612	8.556	8.811	8.631
Num. B's	7.575	8.693	7.862	7.591
Num. C's	7.769	9.019	7.663	8.254
Num. D's	3.591	3.313	2.311	2.555
Num. F's	2.291	2.232	1.144	1.394
Missing Baseline GPA/Grades	0.204	0.029	0.017	0.006
Days Absent	17.404	11.917	9.019	8.567
Missing Attendance Data	0.119	0.000	0.000	0.000
Math Test Score (Z)	-0.572	-0.266	0.235	0.822
Reading Test Score (Z)	-0.523	-0.118	0.264	0.533
Missing Math Test	0.182	0.035	0.036	0.017
Missing Reading Test	0.185	0.039	0.037	0.018
Out-of-School Suspensions	1.487	0.596	0.538	0.149
Disciplinary Incidents	1.177	0.624	0.409	0.221
Any Arrests at Baseline	0.124	0.077	0.073	0.050
Arrests: Violent Crime	0.081	0.047	0.033	0.010
Arrests: Property Crime	0.041	0.024	0.024	0.014
Arrests: Drug Crime	0.019	0.011	0.014	0.003
Math Score - Decile in Previous School	4.097	4.866	6.590	8.446
In Study 2	0.506	0.503	0.454	0.430
Participated in Year 1 of Study	0.289	0.286	0.228	0.250

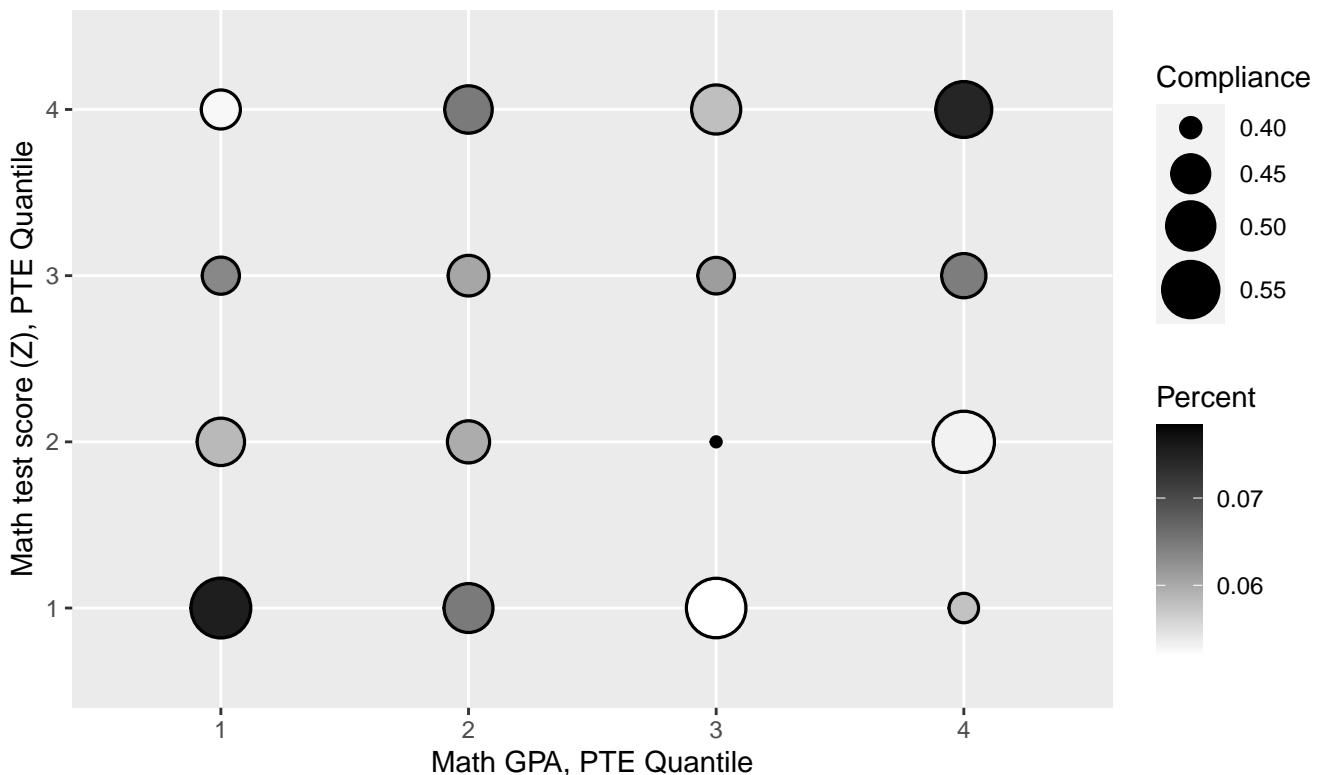
0.3287927 25 1 0.7360488 1 0.003484605 2.304497

## Decile Heatmaps

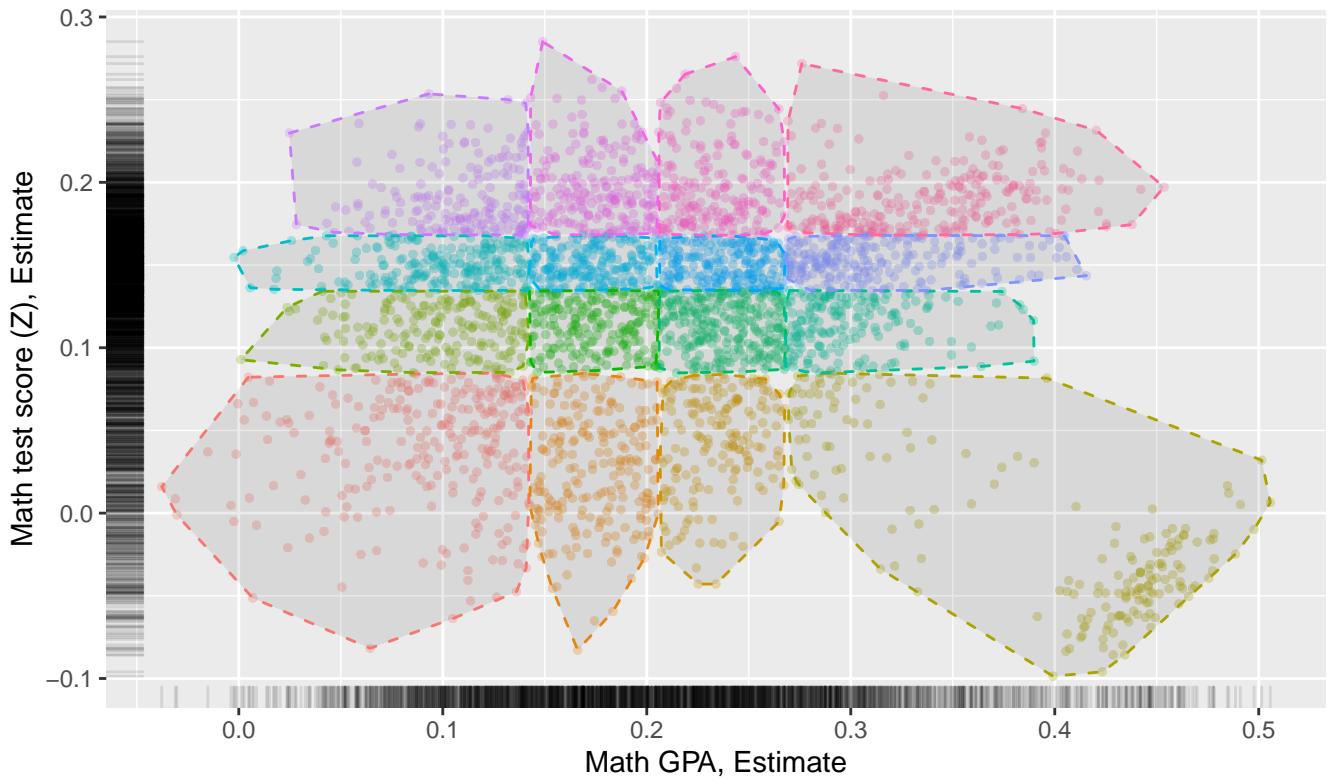
PTE Decile Plot: Math test score (Z) vs. Math GPA



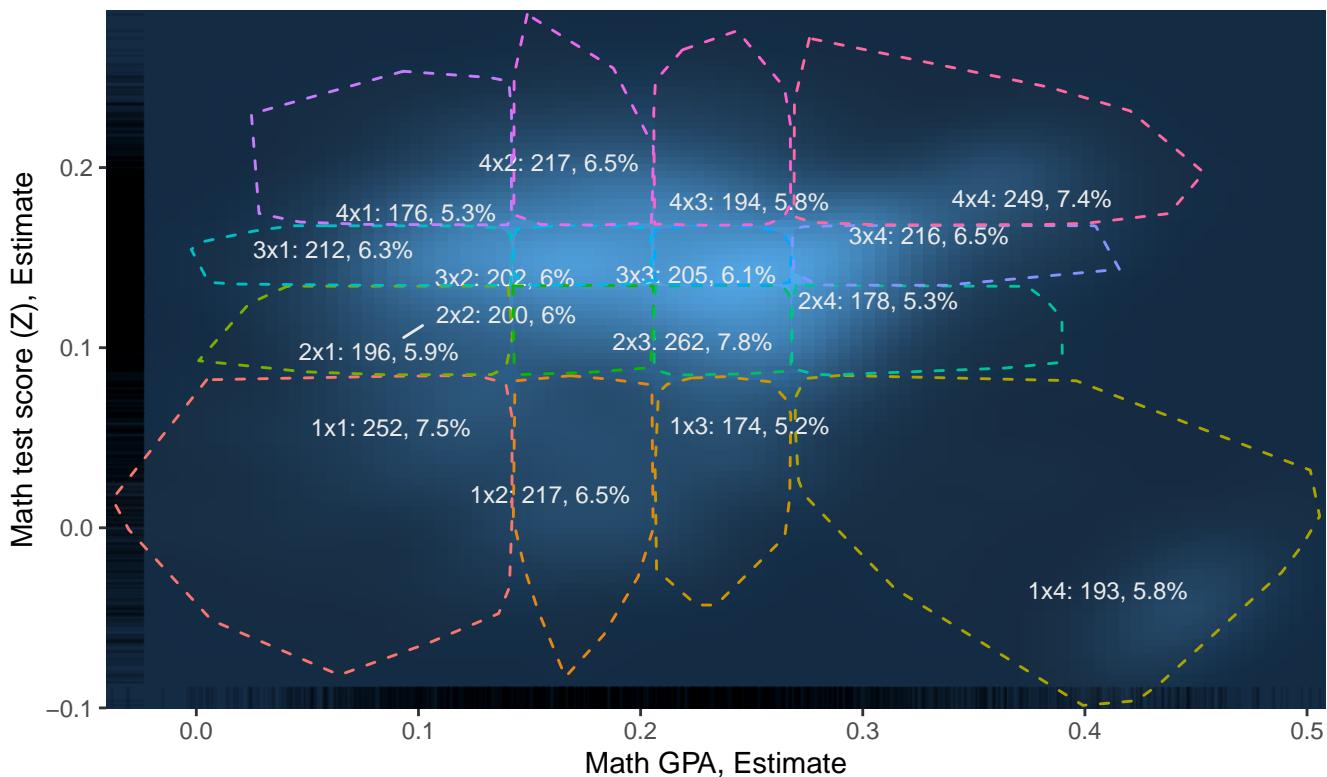
PTE Decile Plot: Math test score (Z) vs. Math GPA



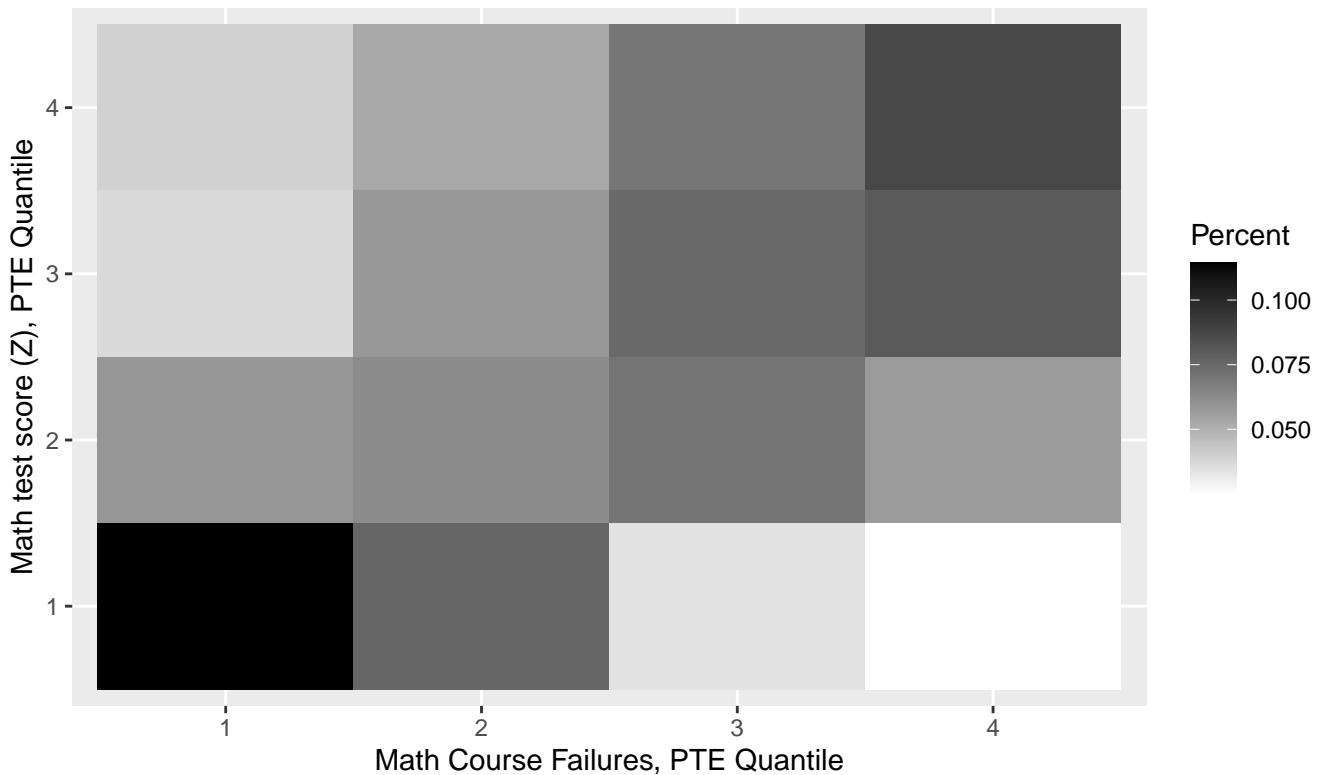
PTE Scatterplot: Math test score (Z) vs Math GPA



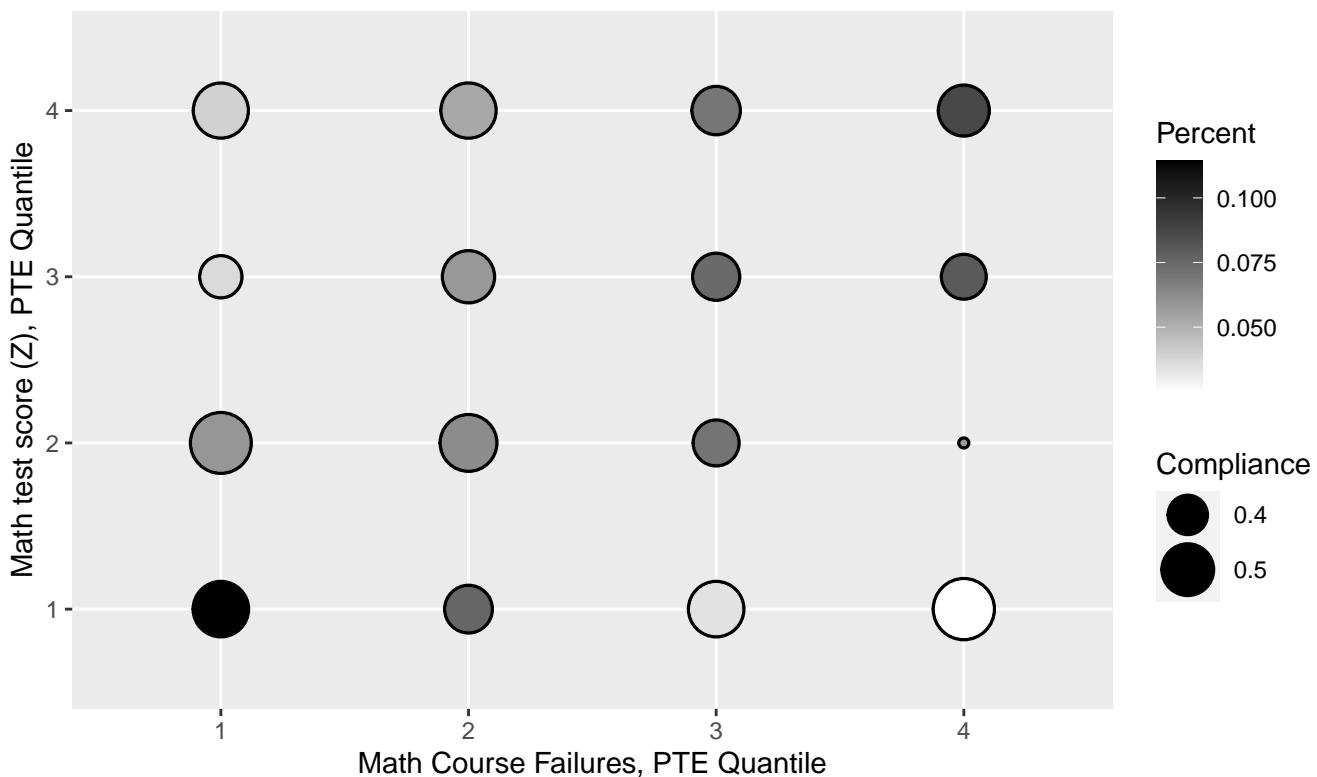
PTE Densities: Math test score (Z) vs Math GPA



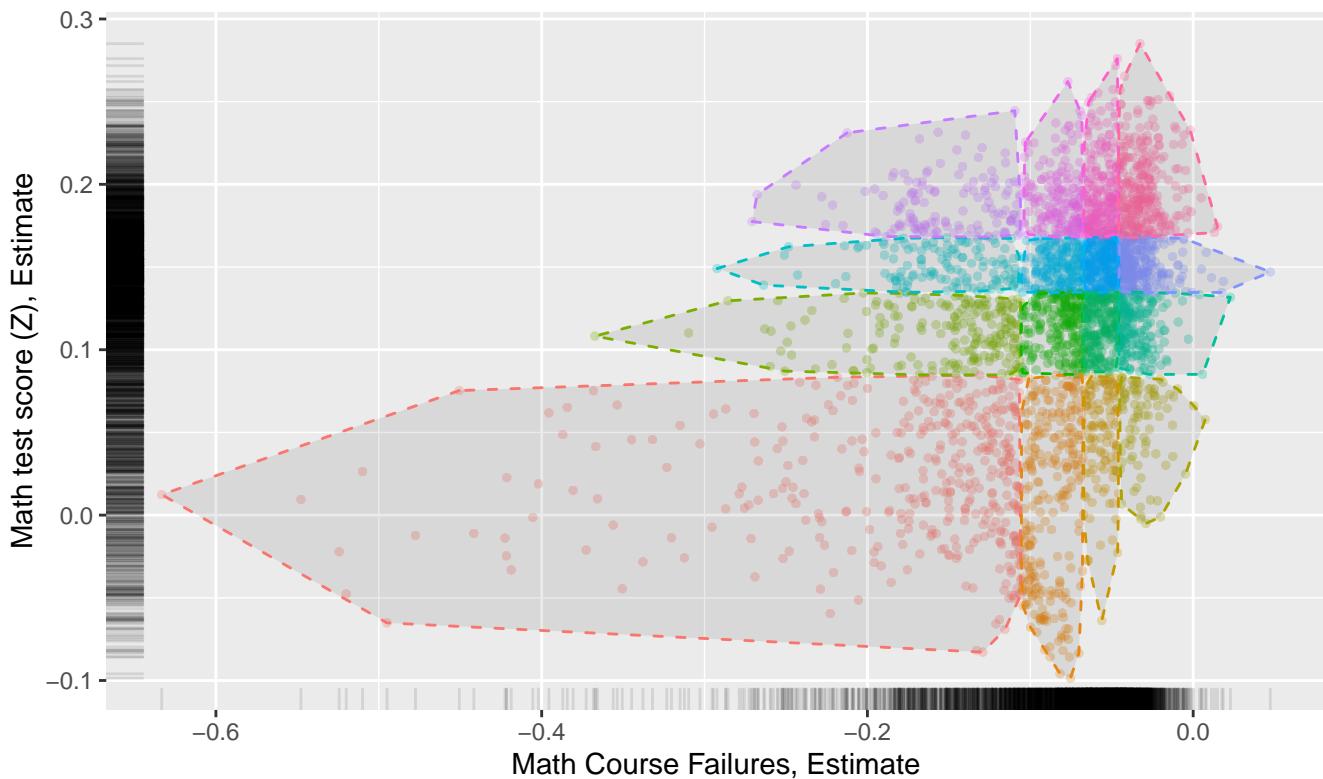
PTE Decile Plot: Math test score (Z) vs. Math Course Failures



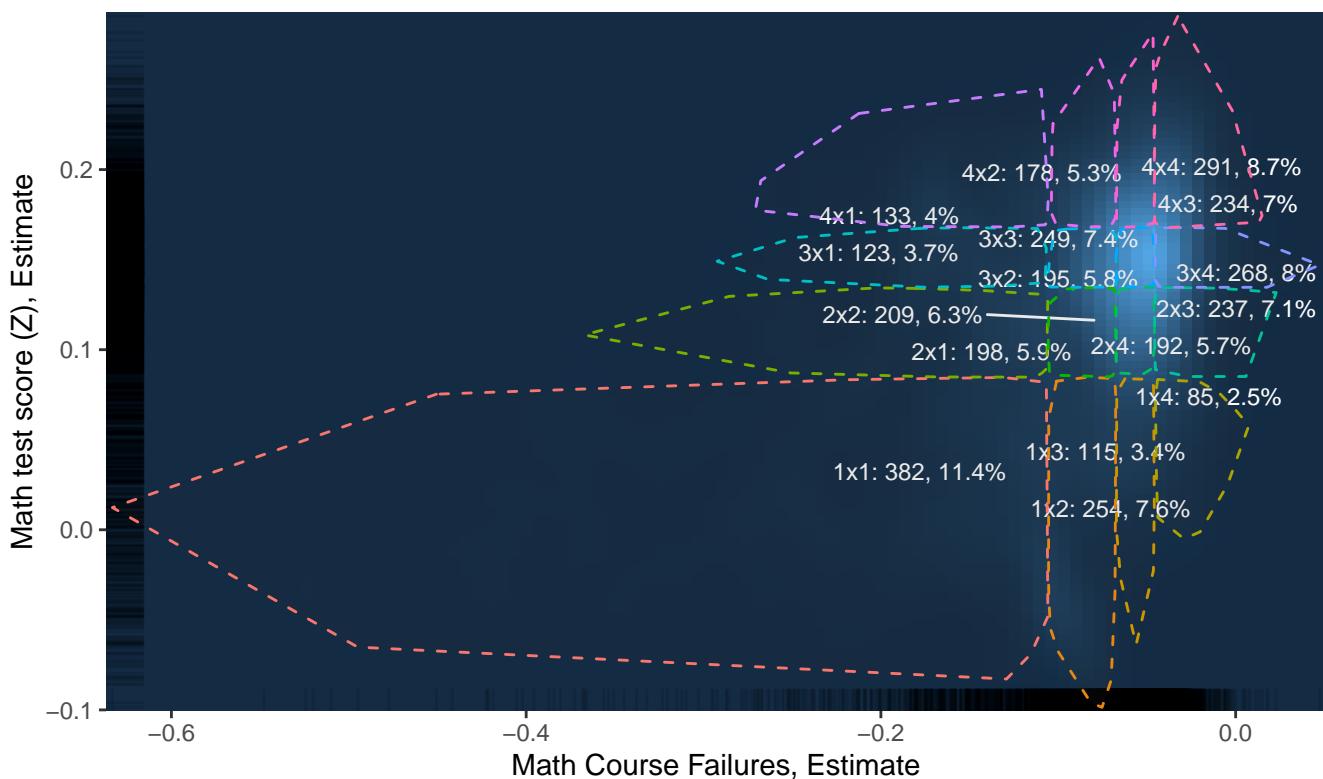
PTE Decile Plot: Math test score (Z) vs. Math Course Failures



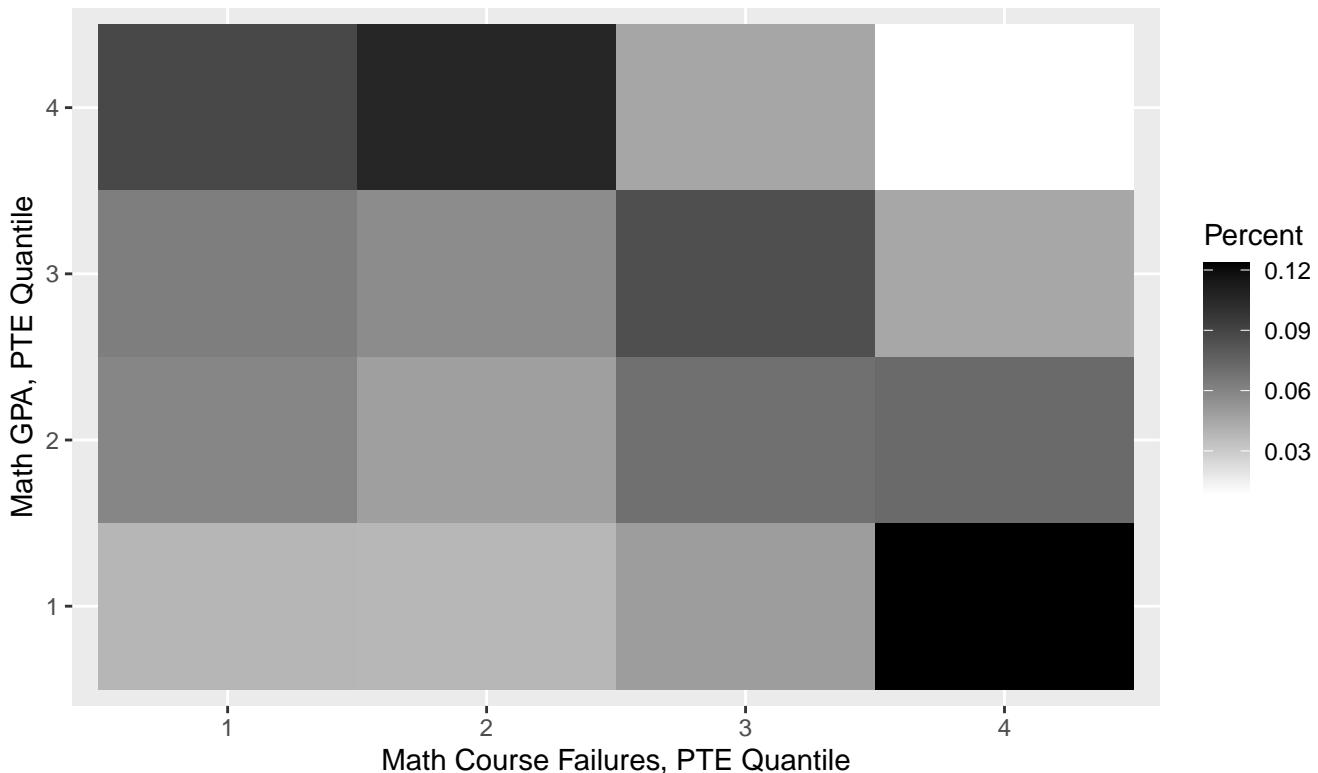
PTE Scatterplot: Math test score (Z) vs Math Course Failures



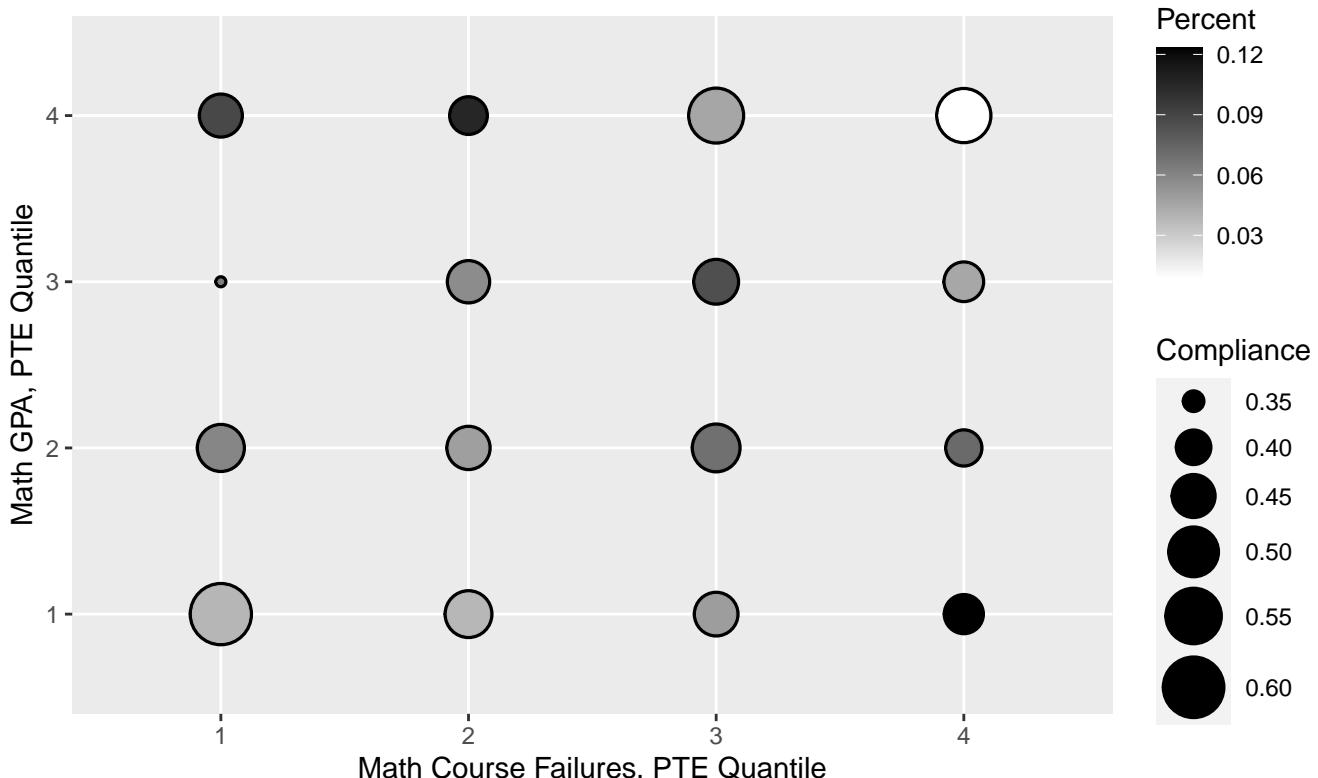
PTE Densities: Math test score (Z) vs Math Course Failures



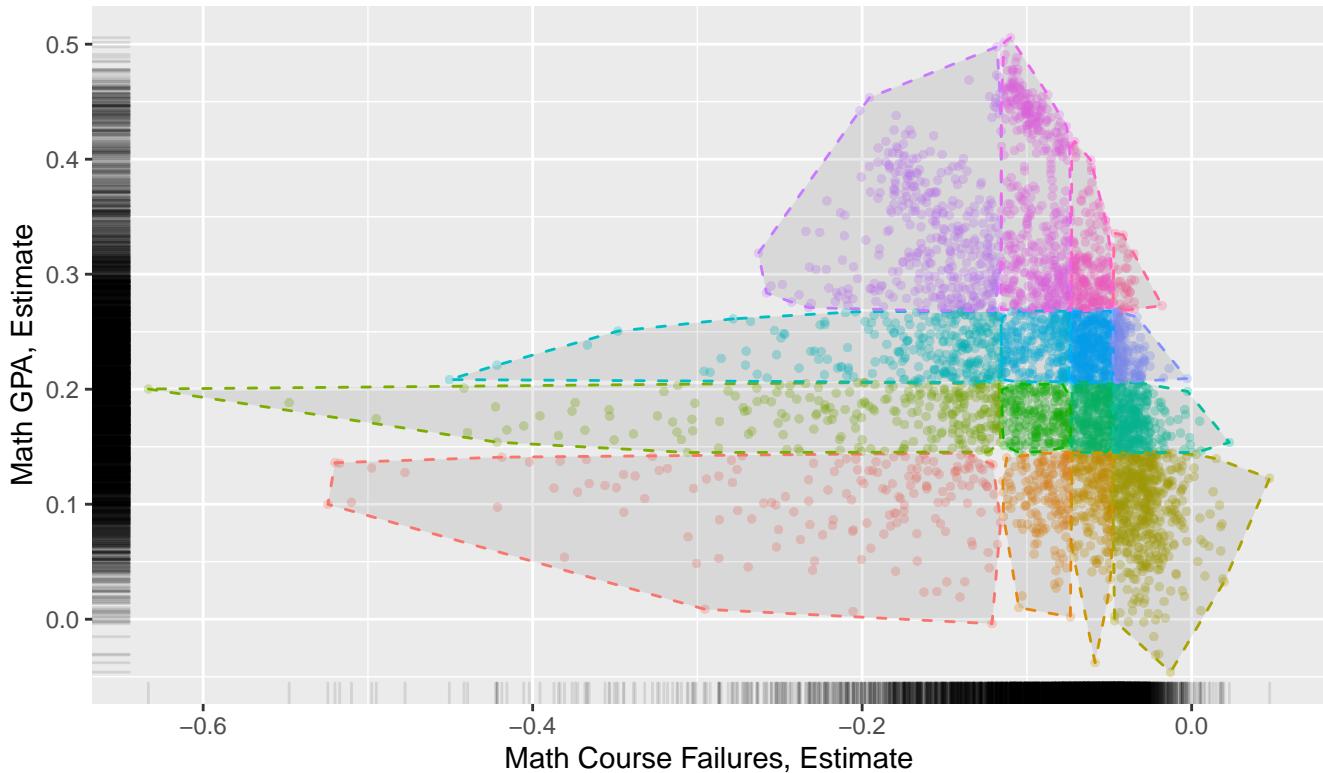
PTE Decile Plot: Math GPA vs. Math Course Failures



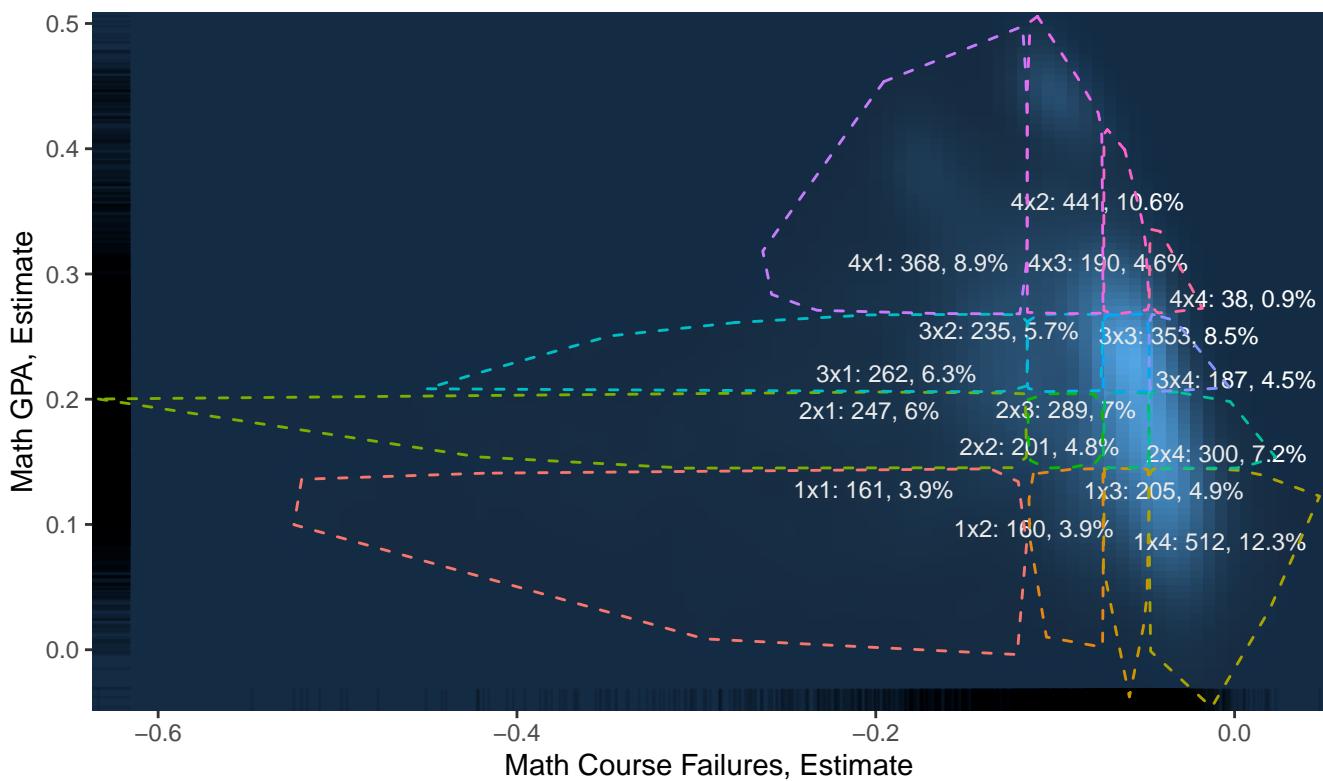
PTE Decile Plot: Math GPA vs. Math Course Failures



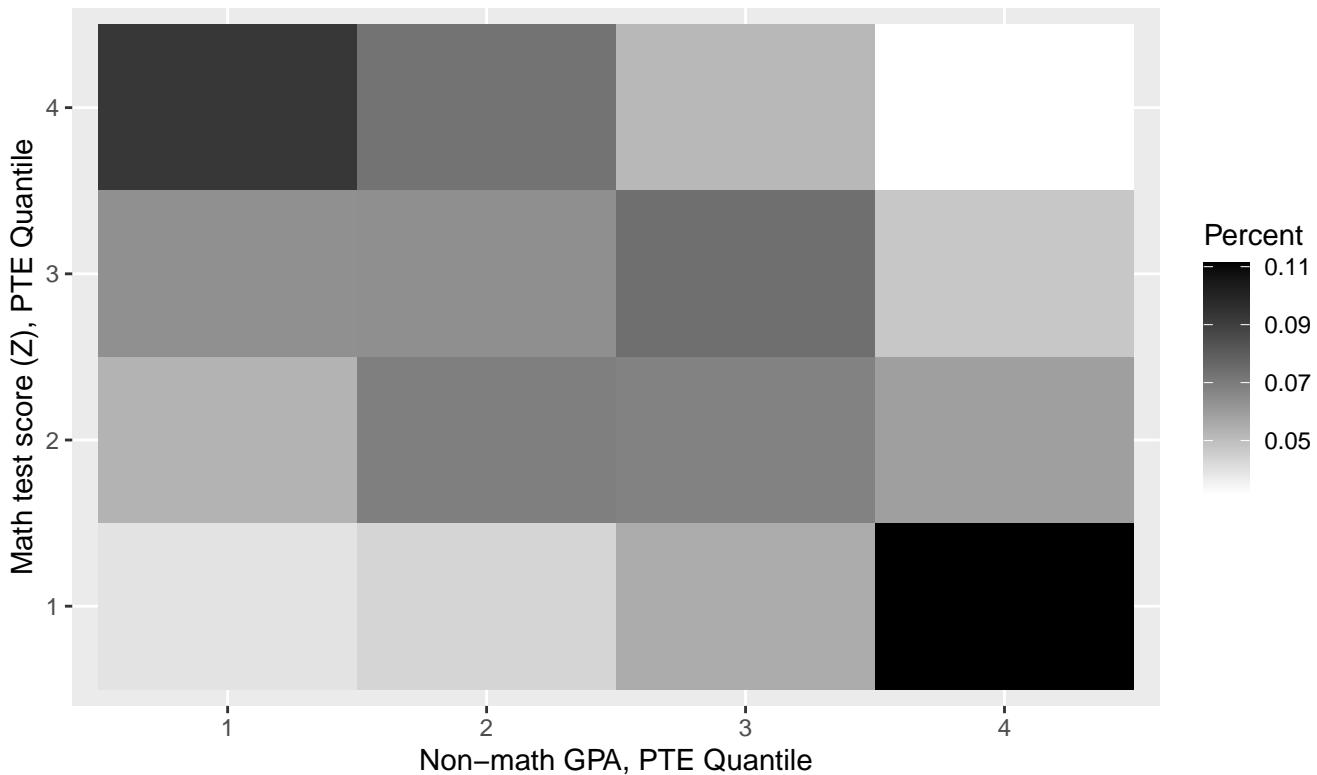
PTE Scatterplot: Math GPA vs Math Course Failures



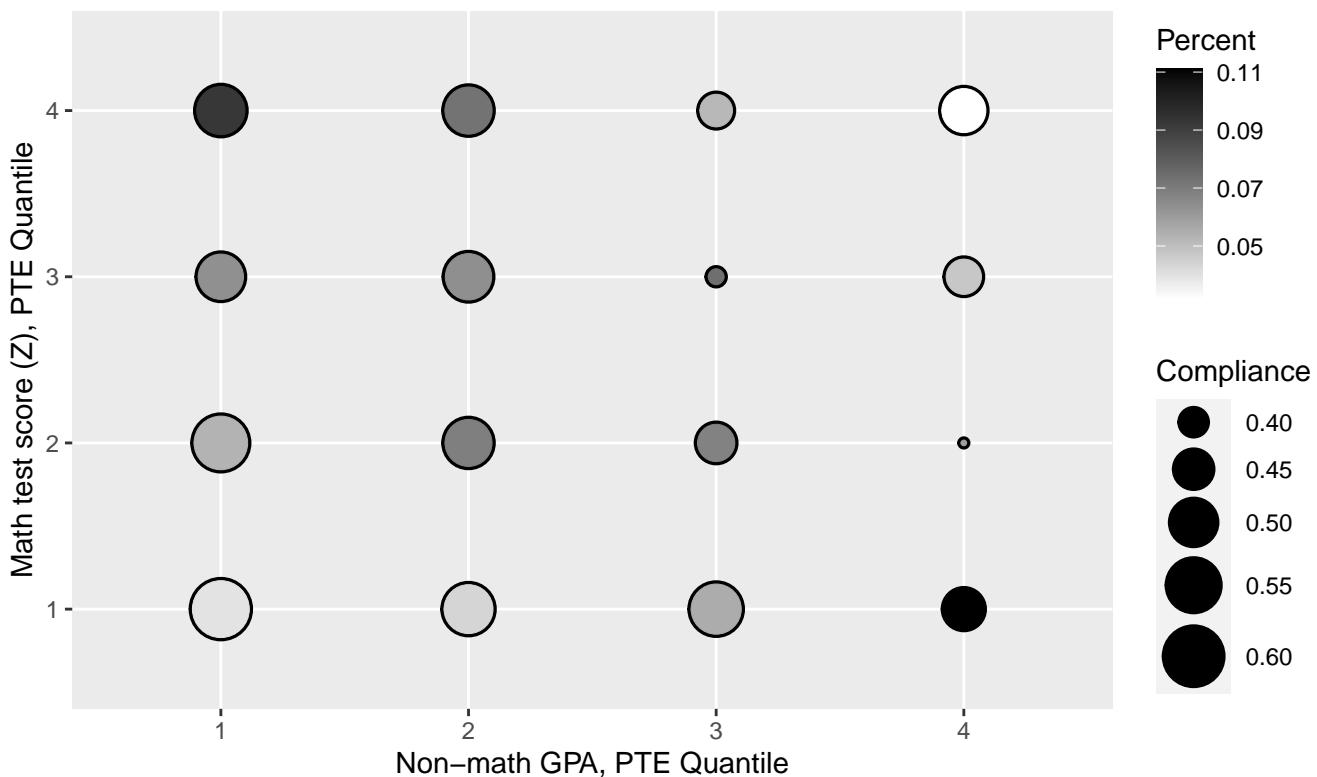
PTE Densities: Math GPA vs Math Course Failures



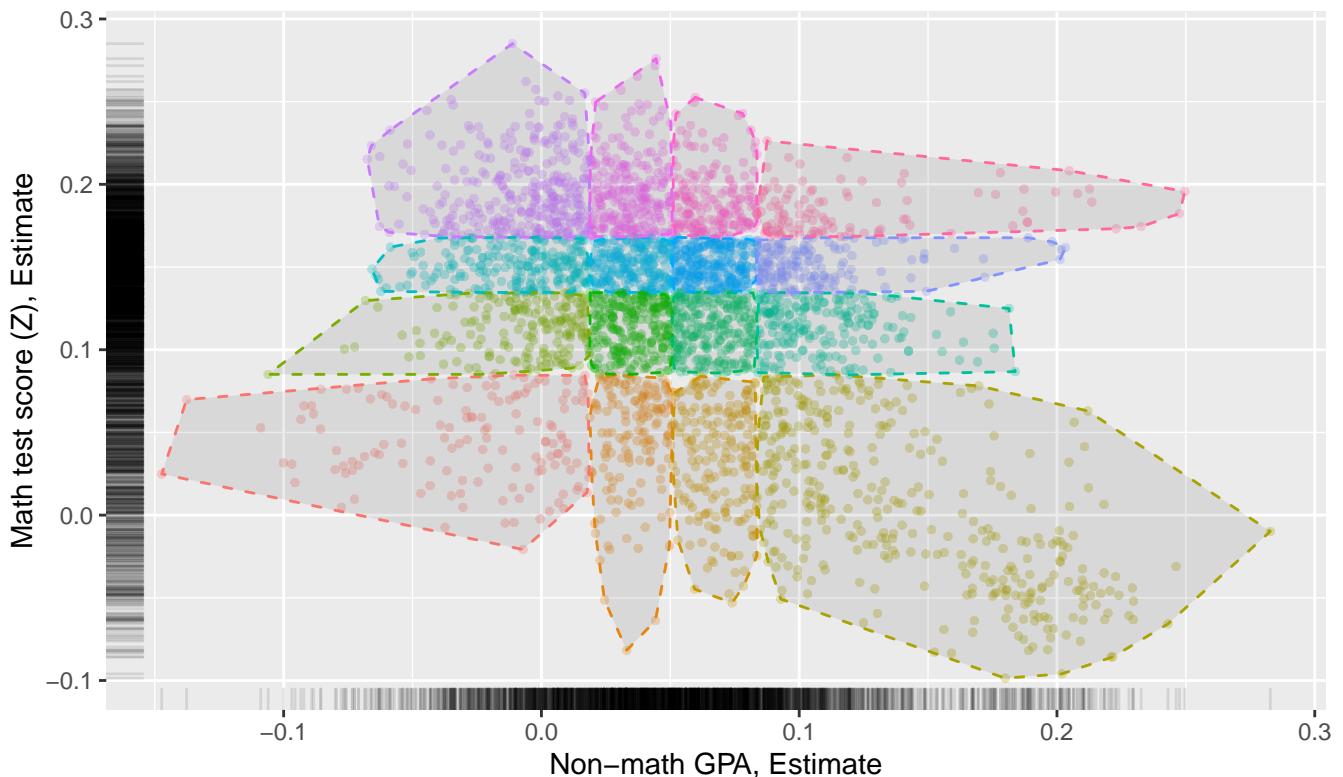
PTE Decile Plot: Math test score (Z) vs. Non–math GPA



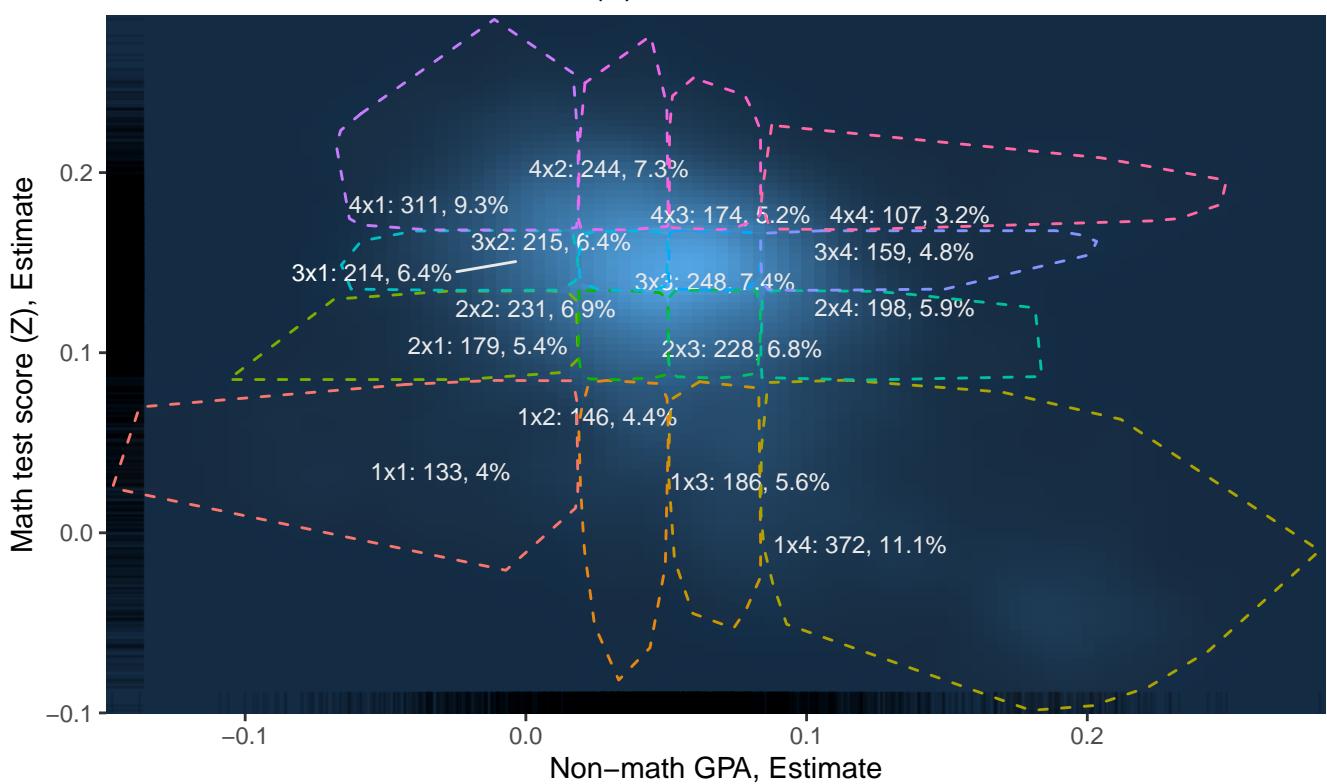
PTE Decile Plot: Math test score (Z) vs. Non–math GPA



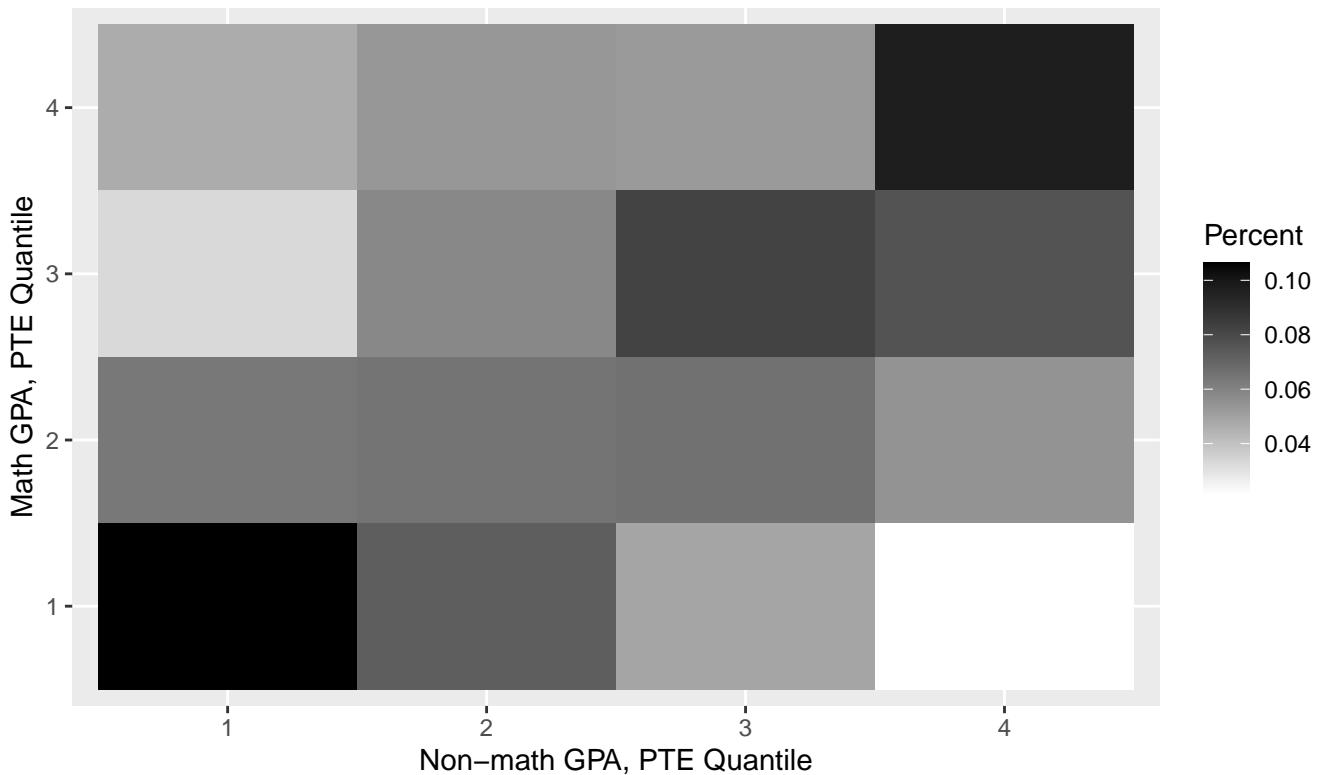
PTE Scatterplot: Math test score (Z) vs Non–math GPA



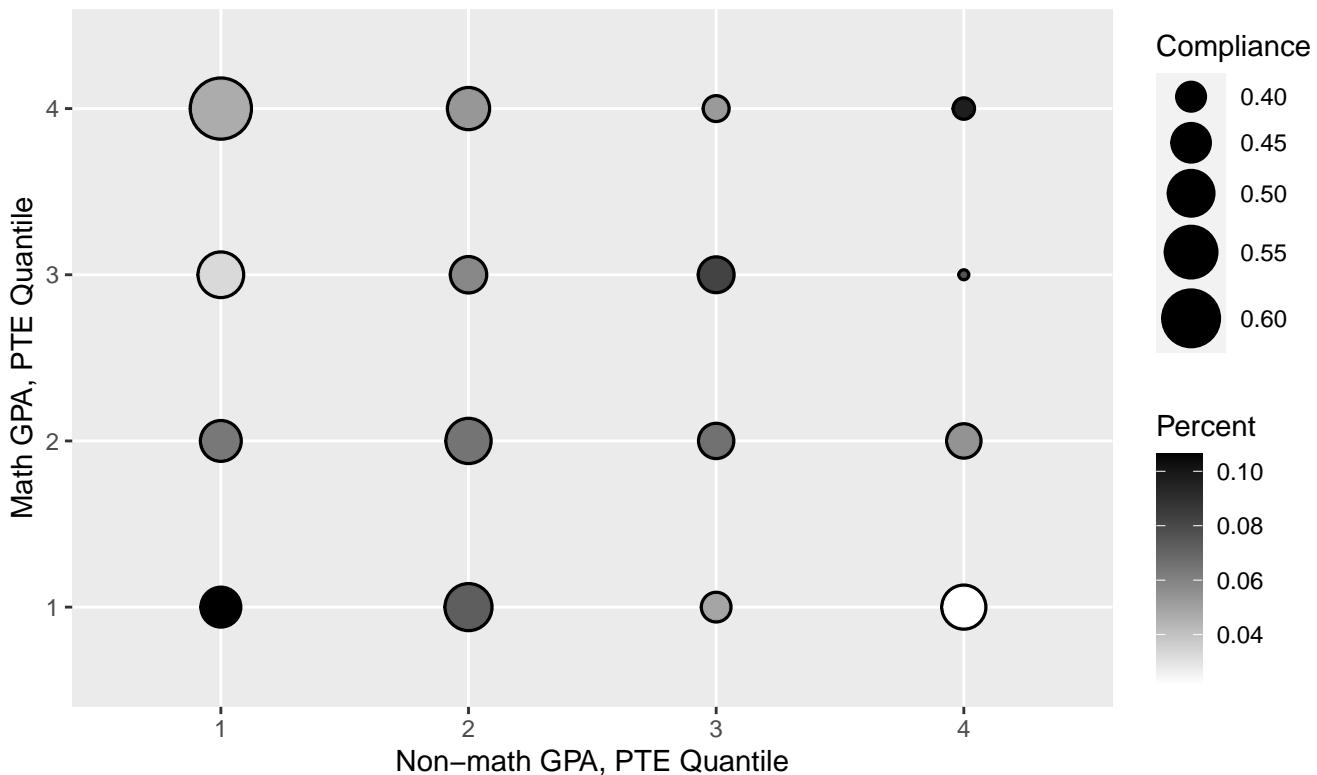
PTE Densities: Math test score (Z) vs Non–math GPA



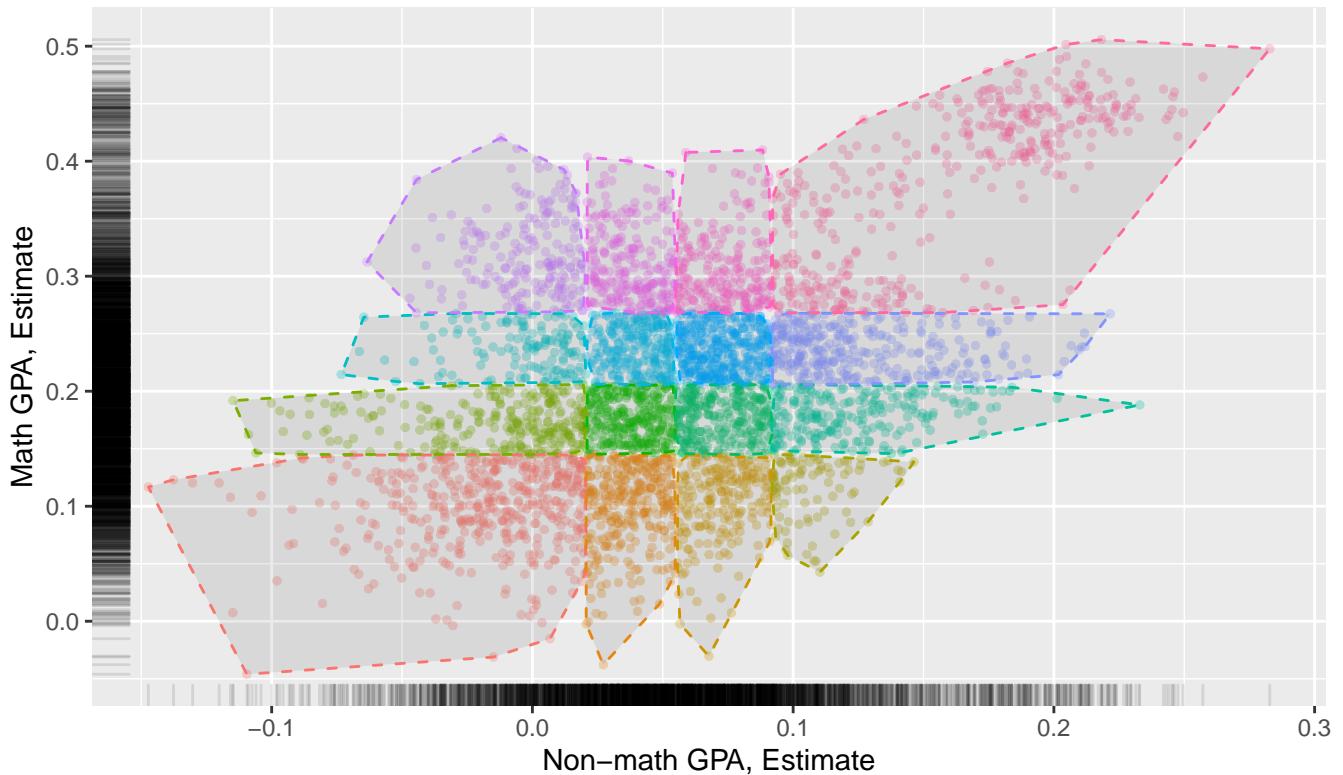
PTE Decile Plot: Math GPA vs. Non–math GPA



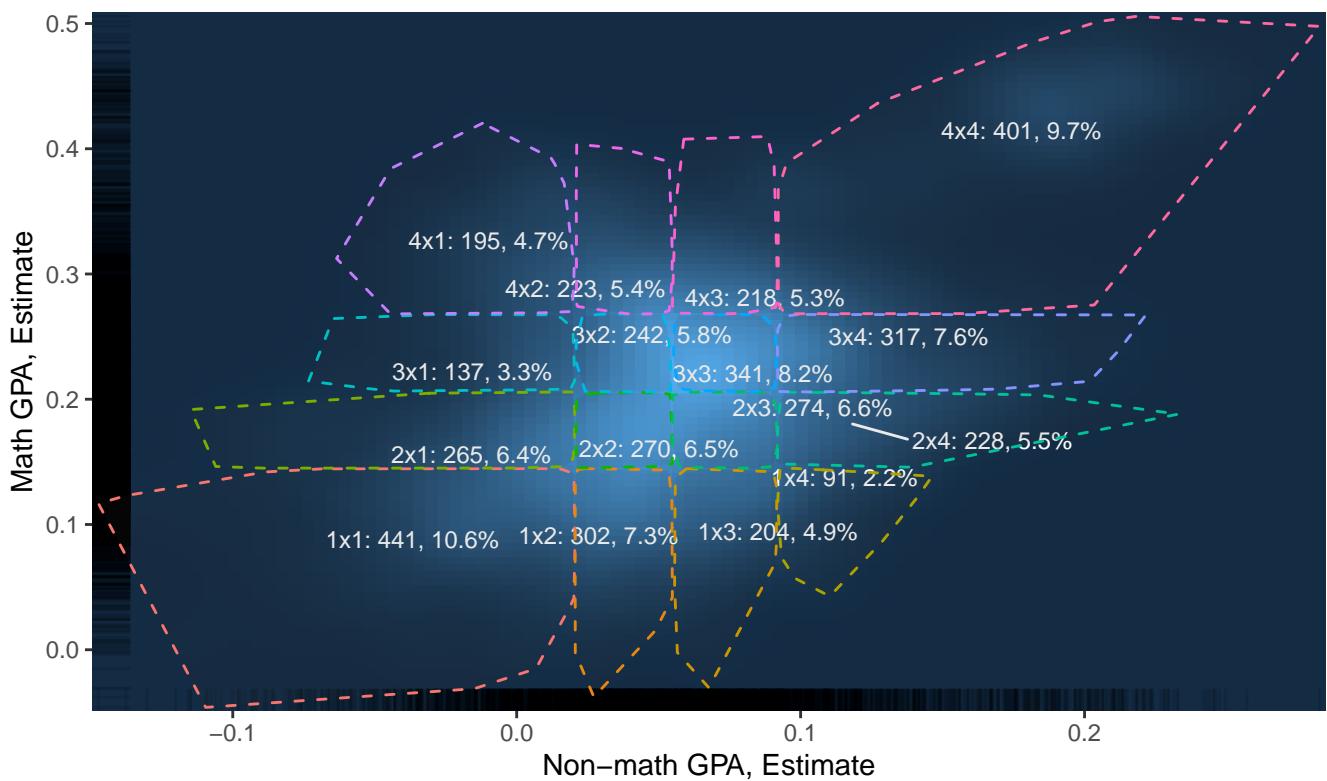
PTE Decile Plot: Math GPA vs. Non–math GPA



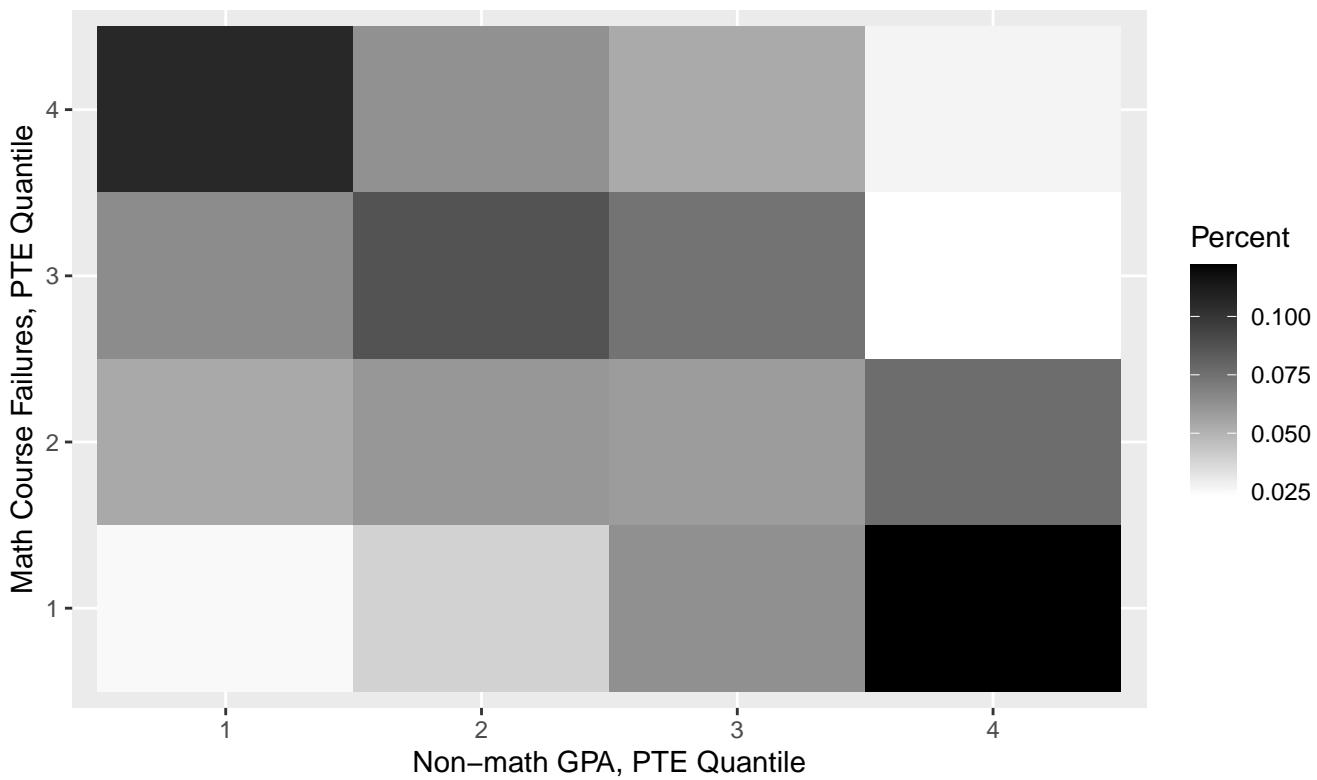
PTE Scatterplot: Math GPA vs Non–math GPA



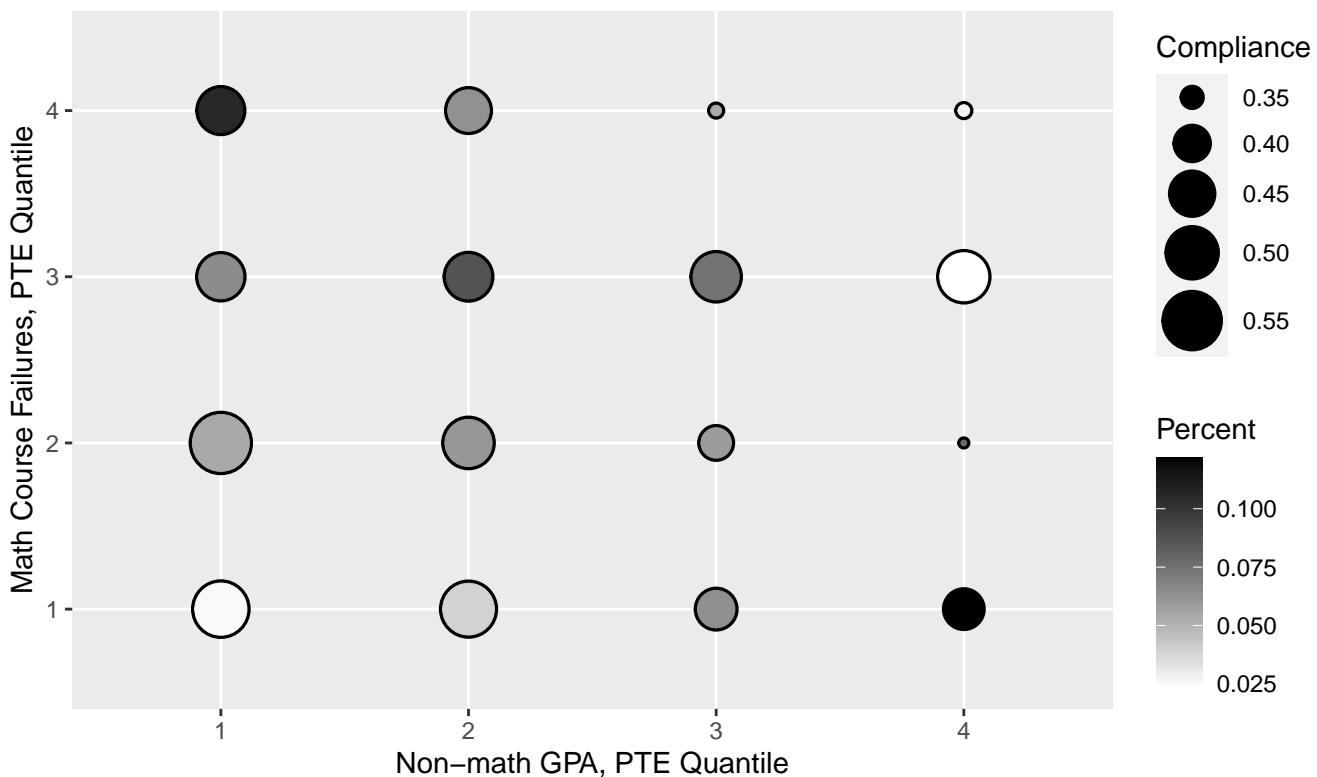
PTE Densities: Math GPA vs Non–math GPA



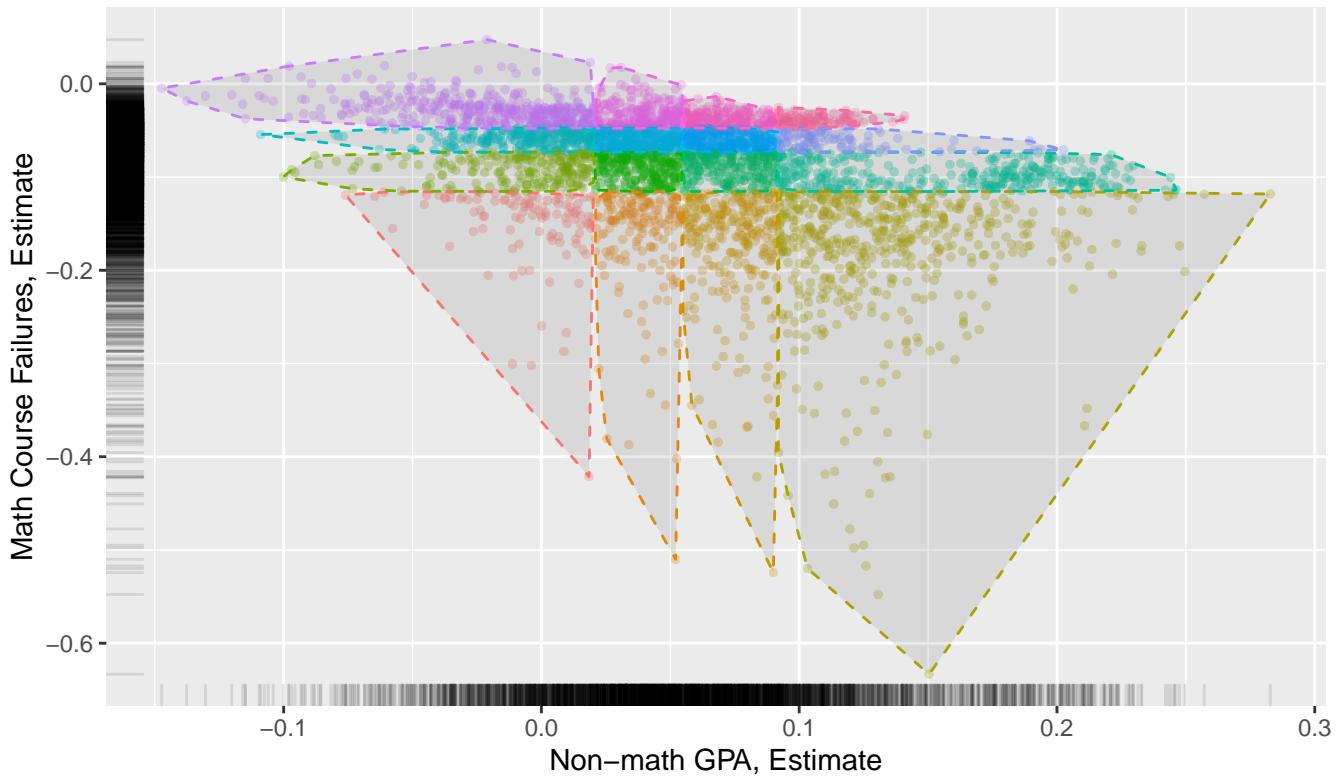
PTE Decile Plot: Math Course Failures vs. Non–math GPA



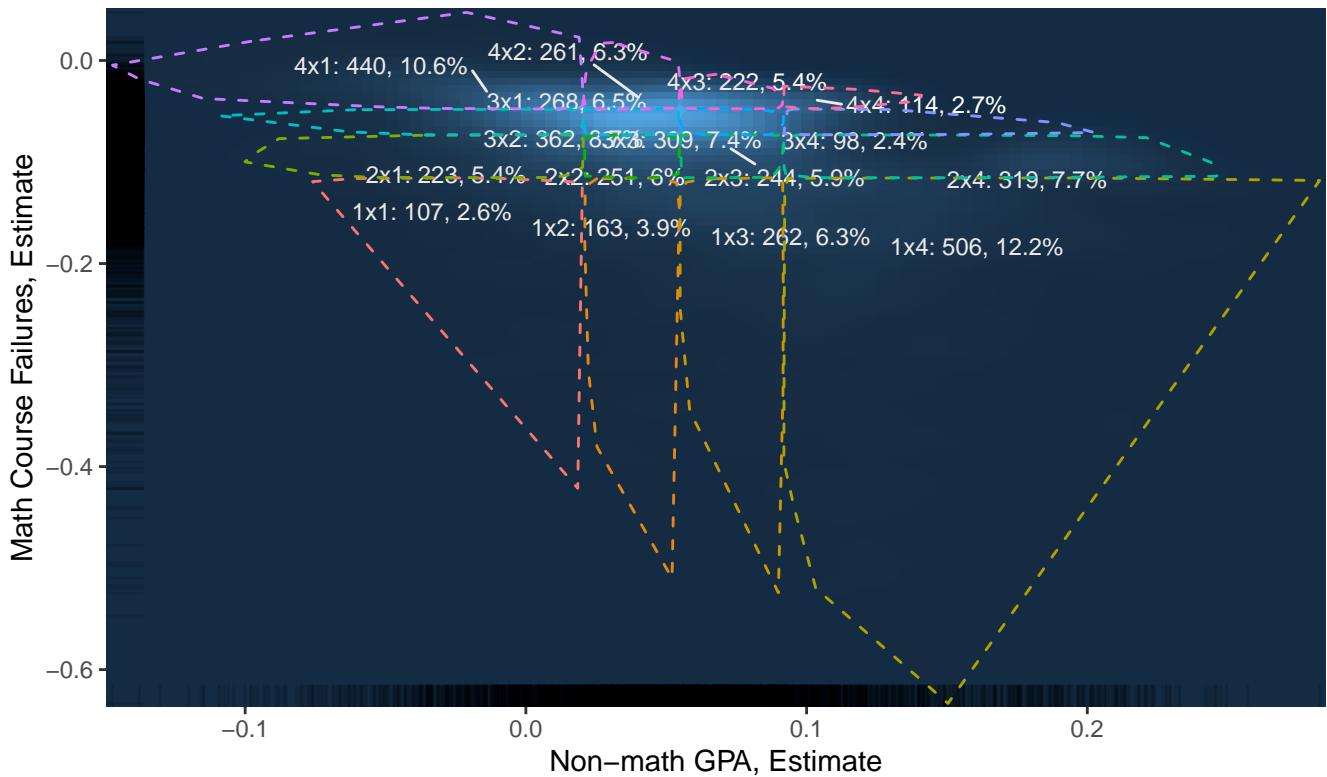
PTE Decile Plot: Math Course Failures vs. Non–math GPA



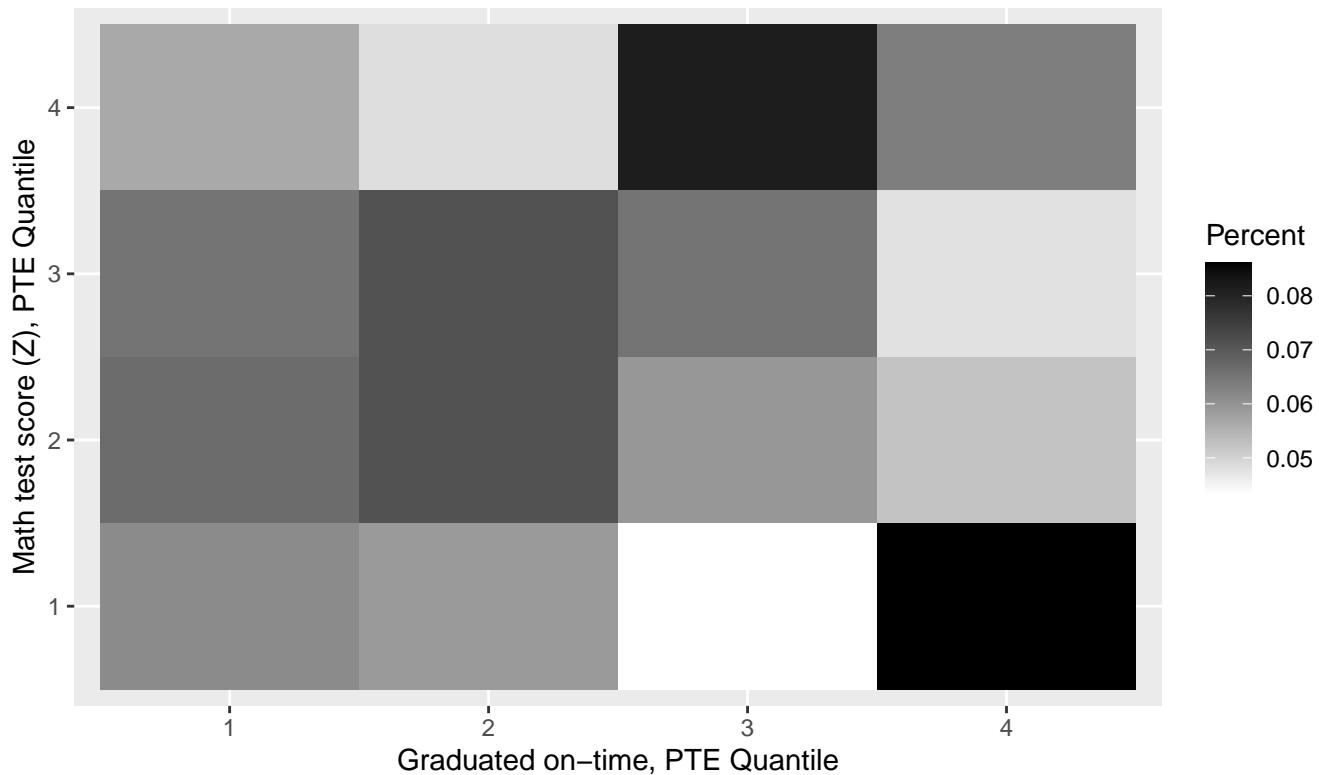
PTE Scatterplot: Math Course Failures vs Non–math GPA



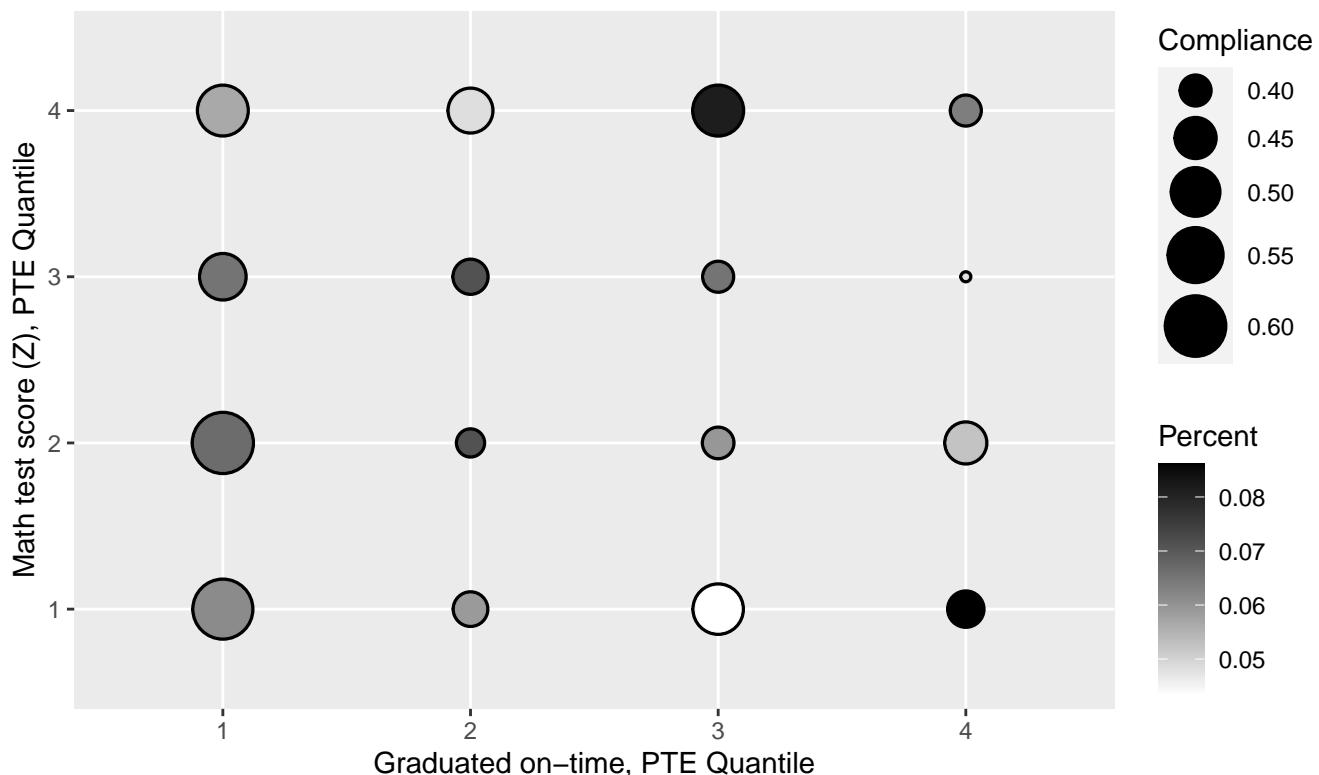
PTE Densities: Math Course Failures vs Non–math GPA



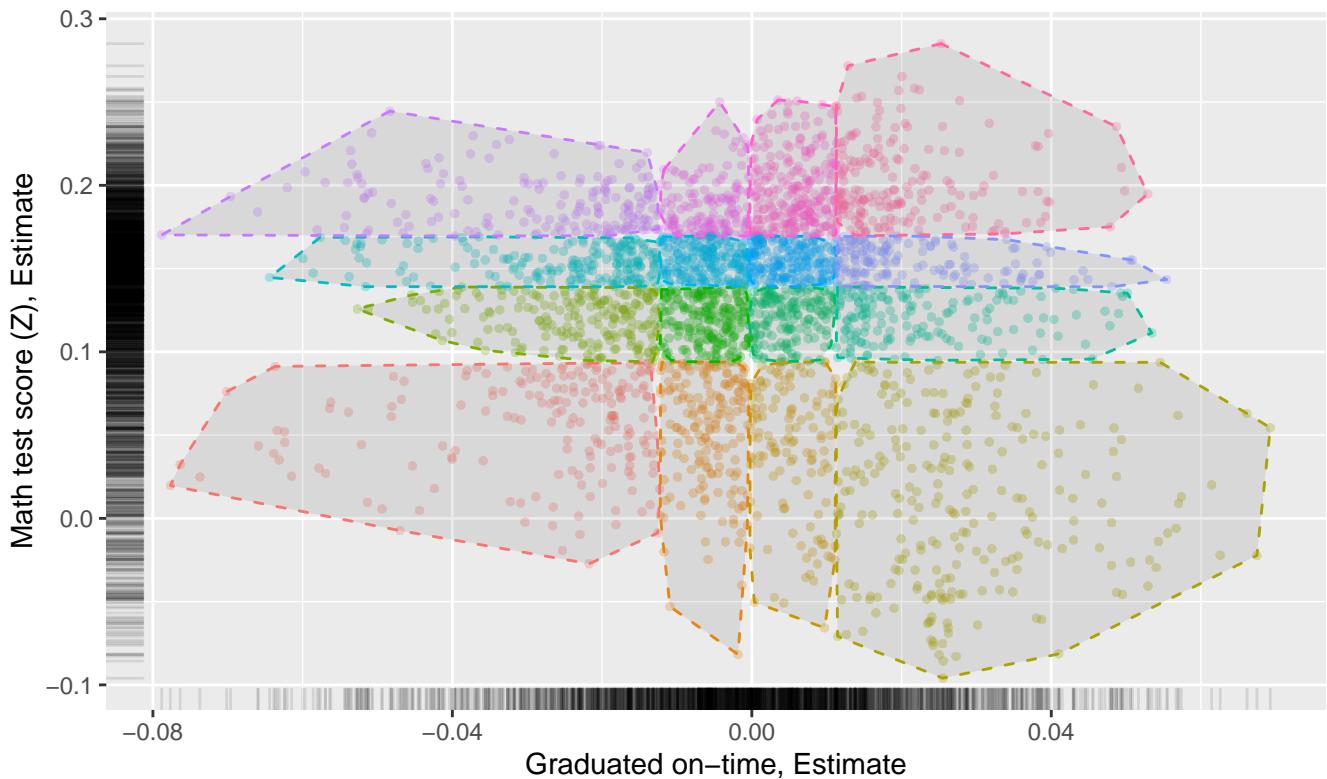
PTE Decile Plot: Math test score (Z) vs. Graduated on-time



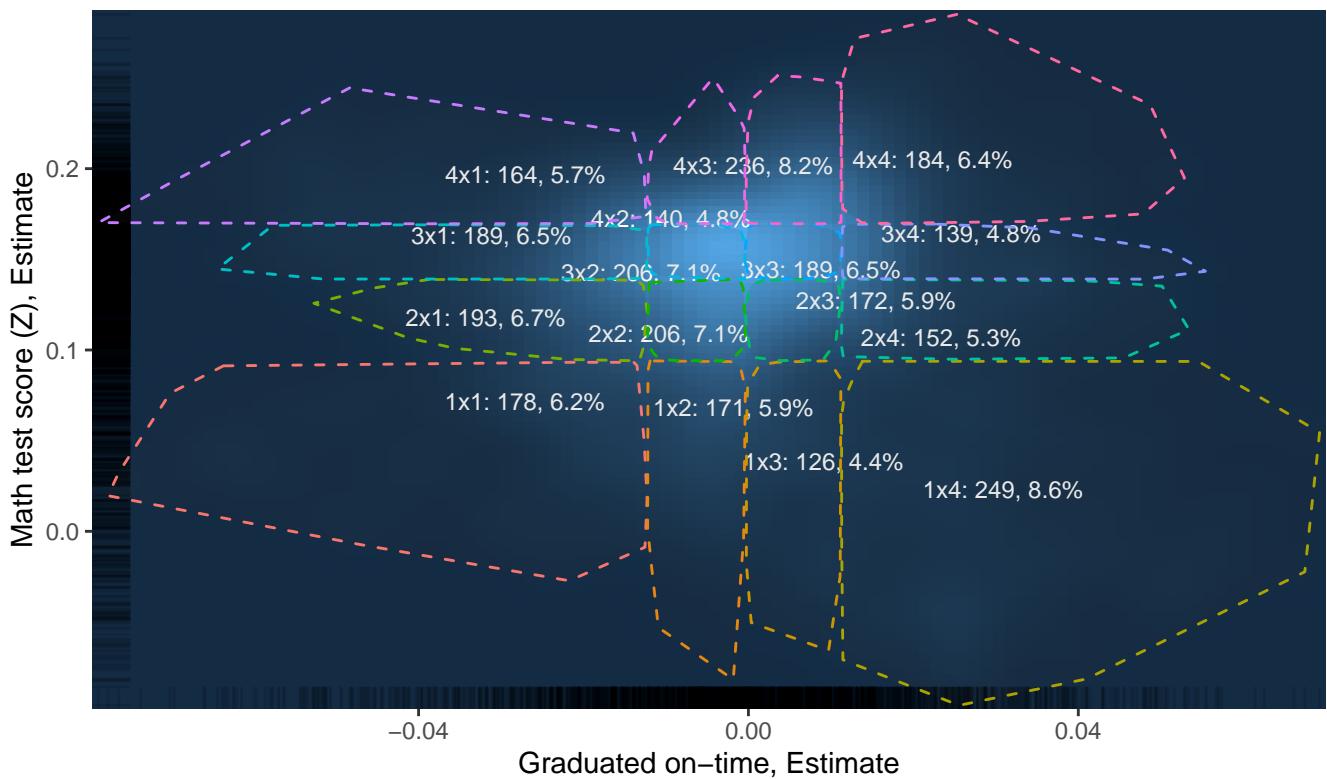
PTE Decile Plot: Math test score (Z) vs. Graduated on-time



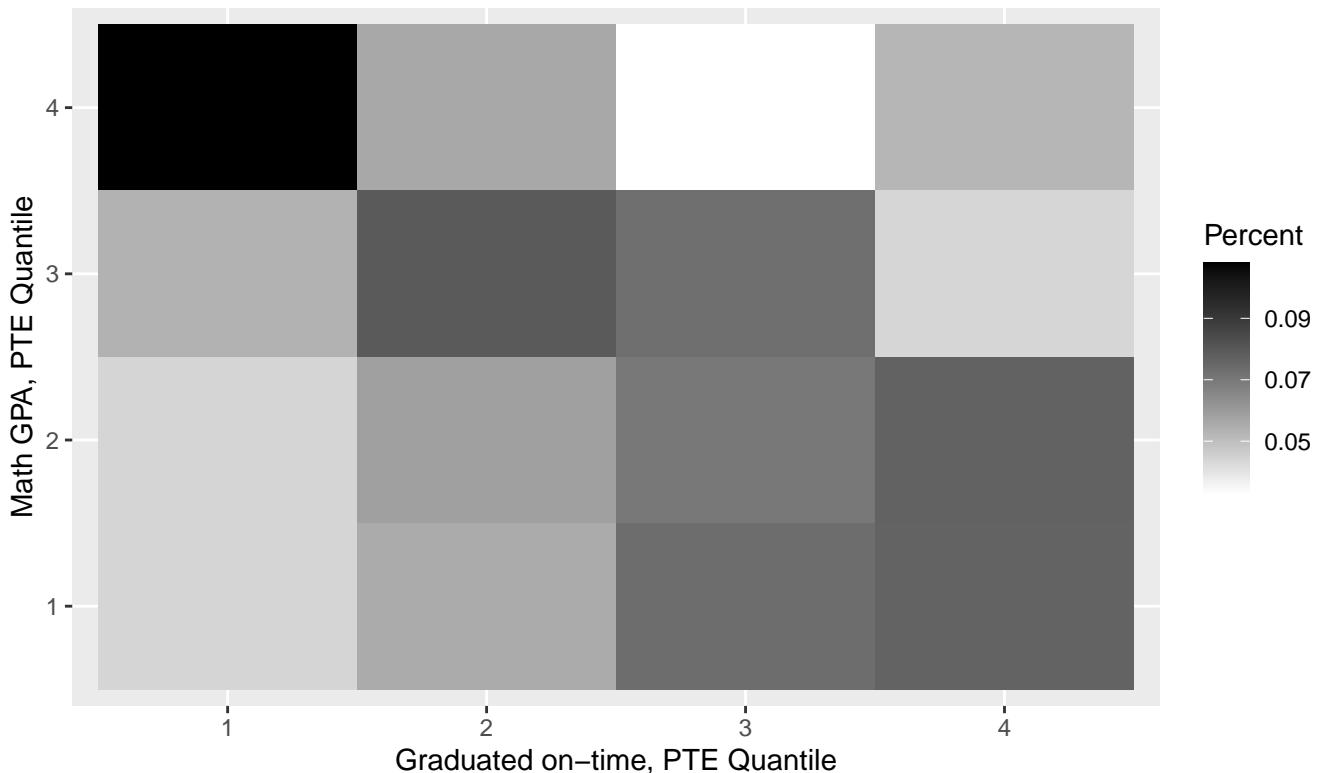
PTE Scatterplot: Math test score (Z) vs Graduated on-time



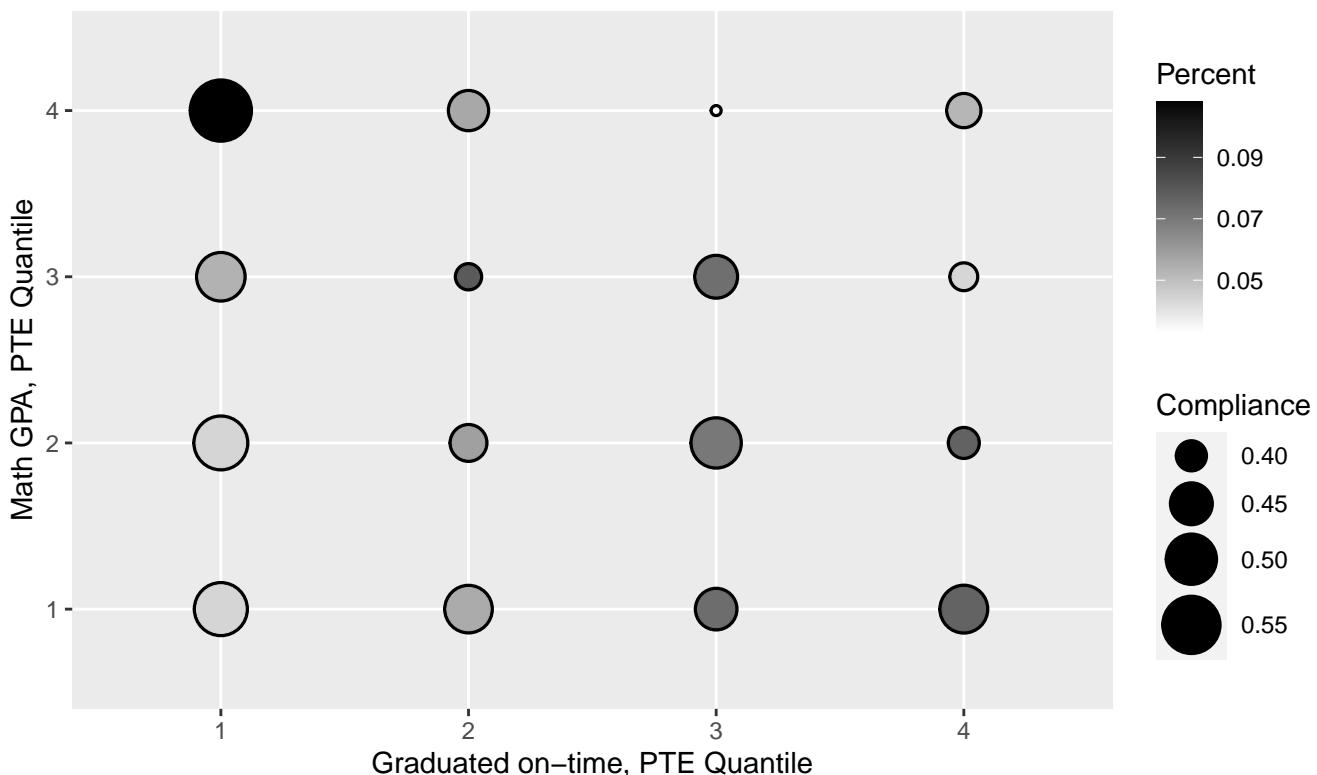
PTE Densities: Math test score (Z) vs Graduated on-time



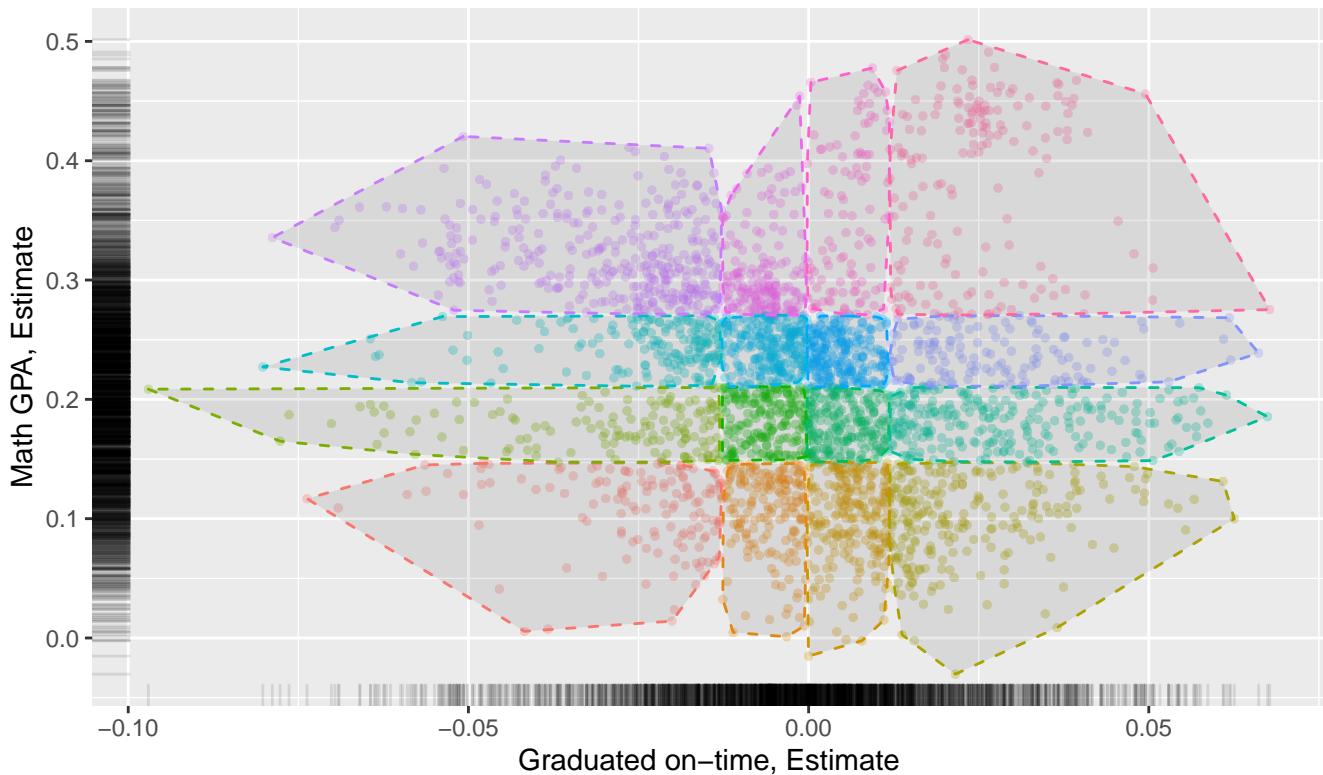
PTE Decile Plot: Math GPA vs. Graduated on-time



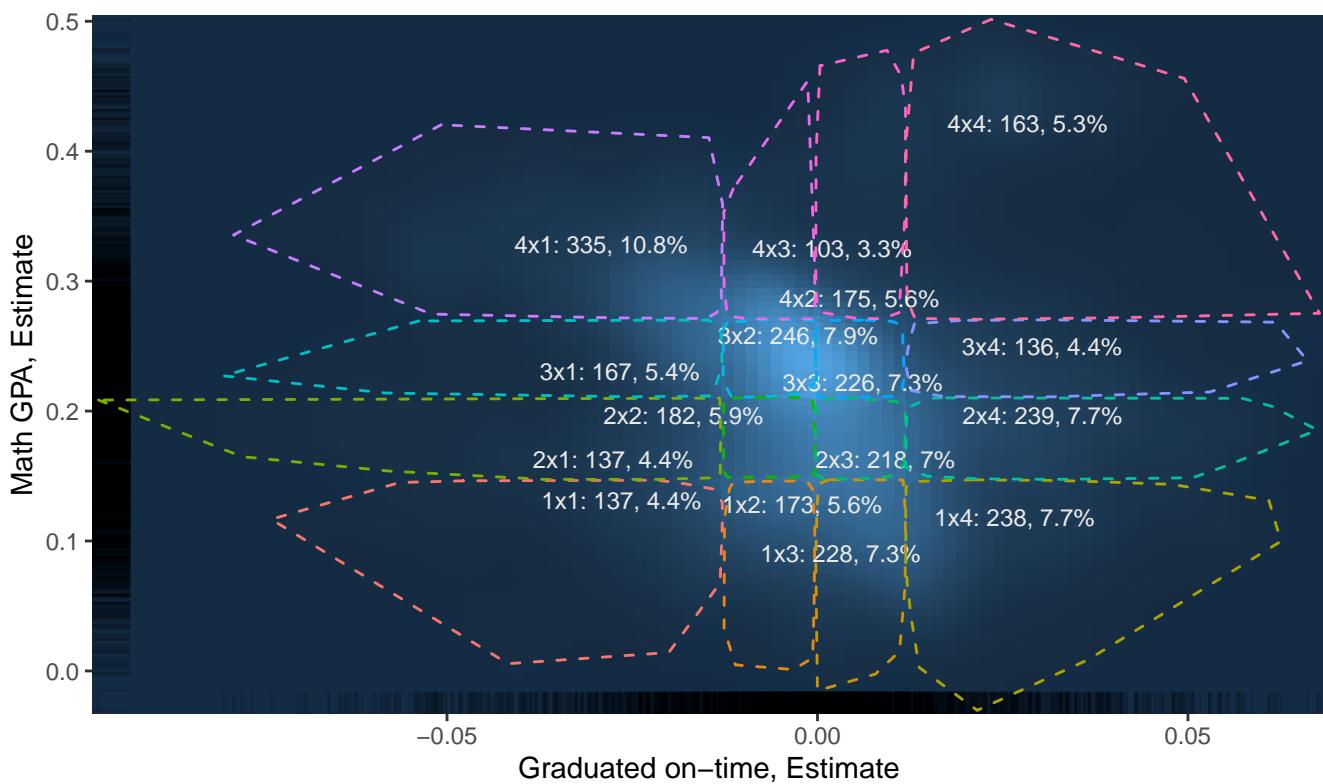
PTE Decile Plot: Math GPA vs. Graduated on-time



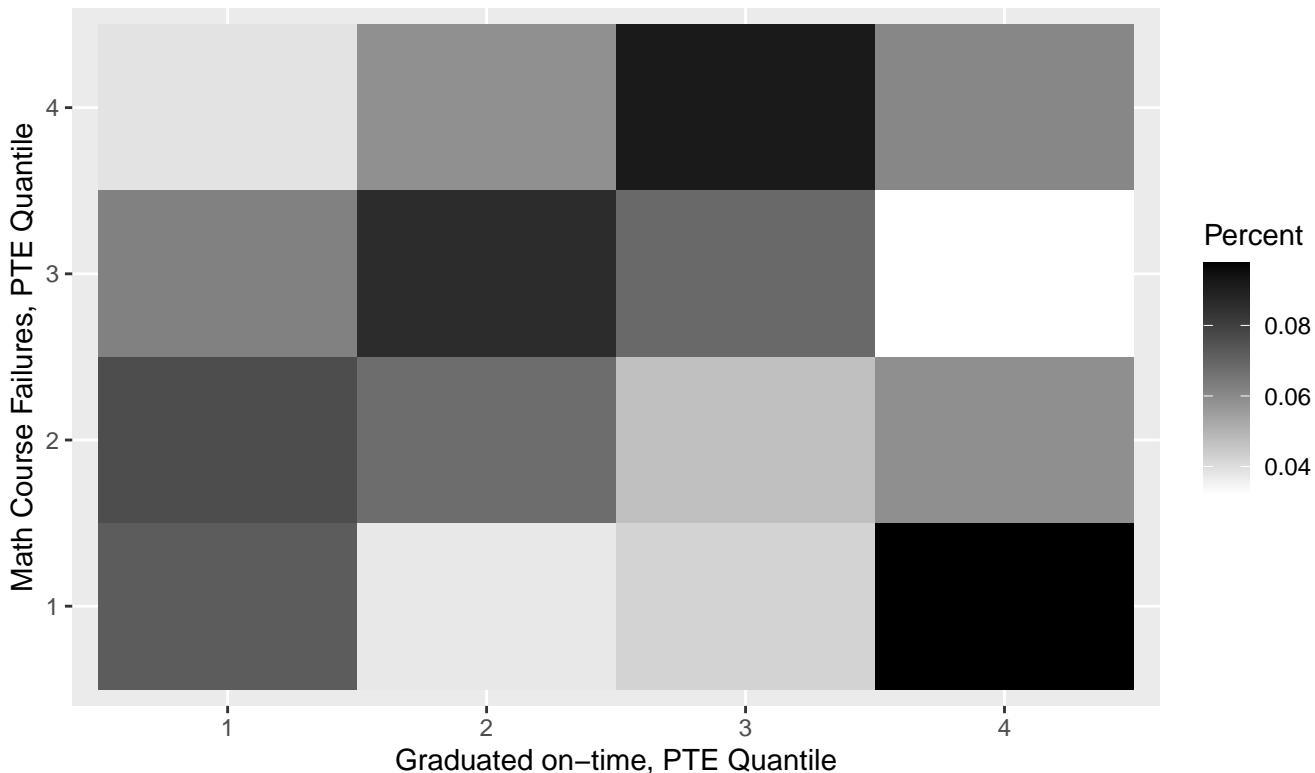
PTE Scatterplot: Math GPA vs Graduated on-time



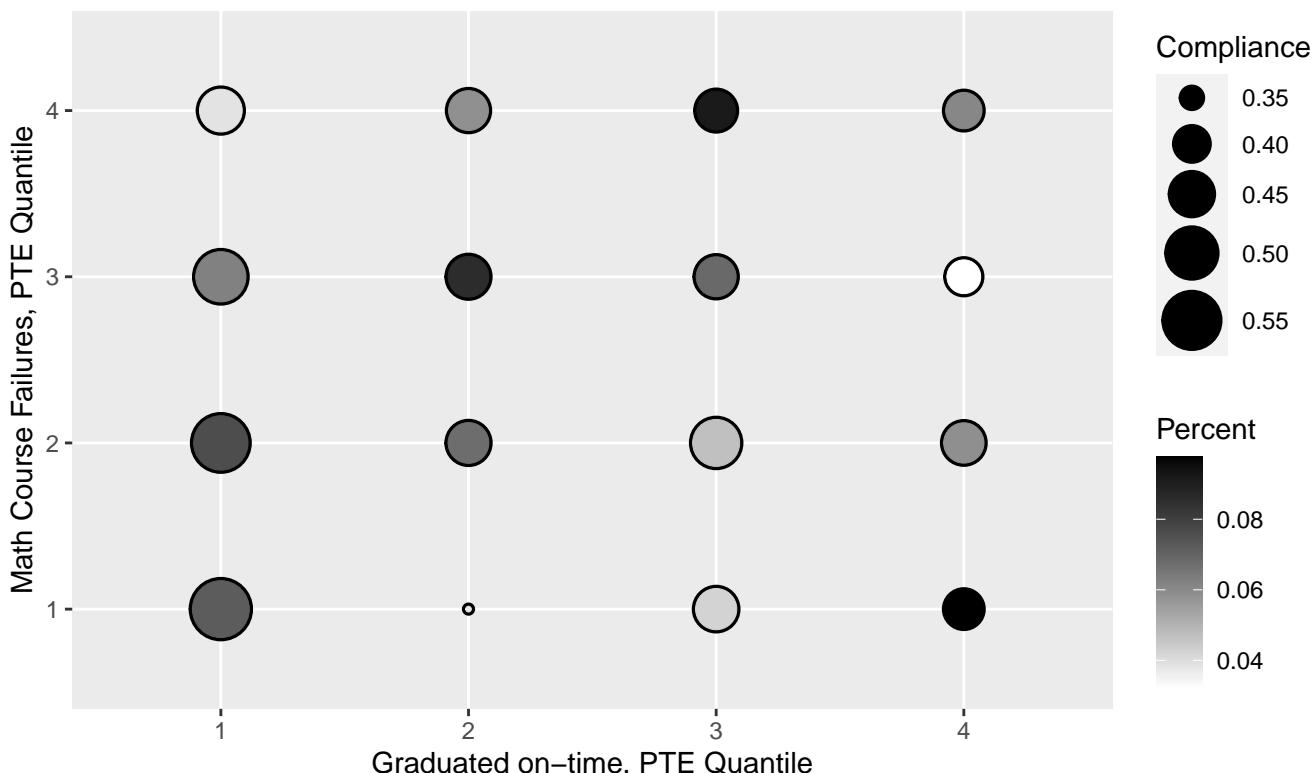
PTE Densities: Math GPA vs Graduated on-time



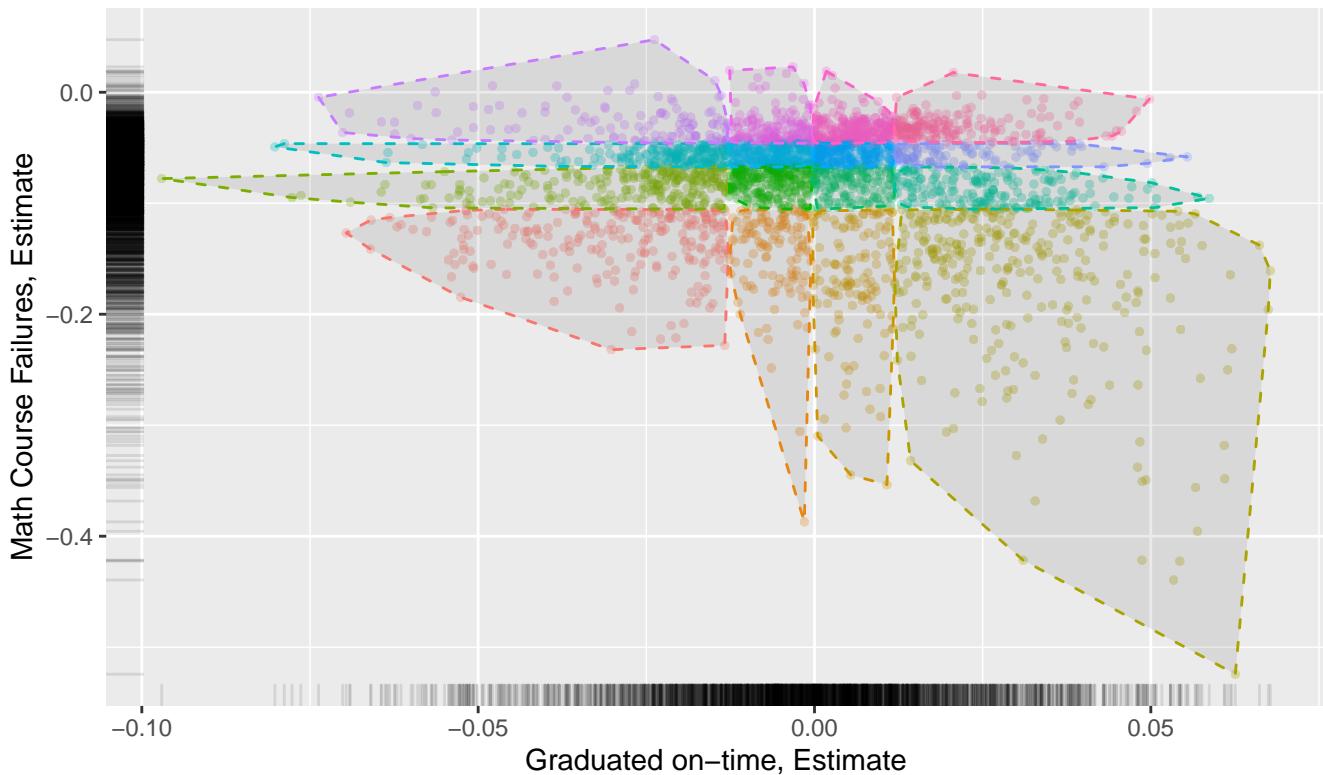
PTE Decile Plot: Math Course Failures vs. Graduated on-time



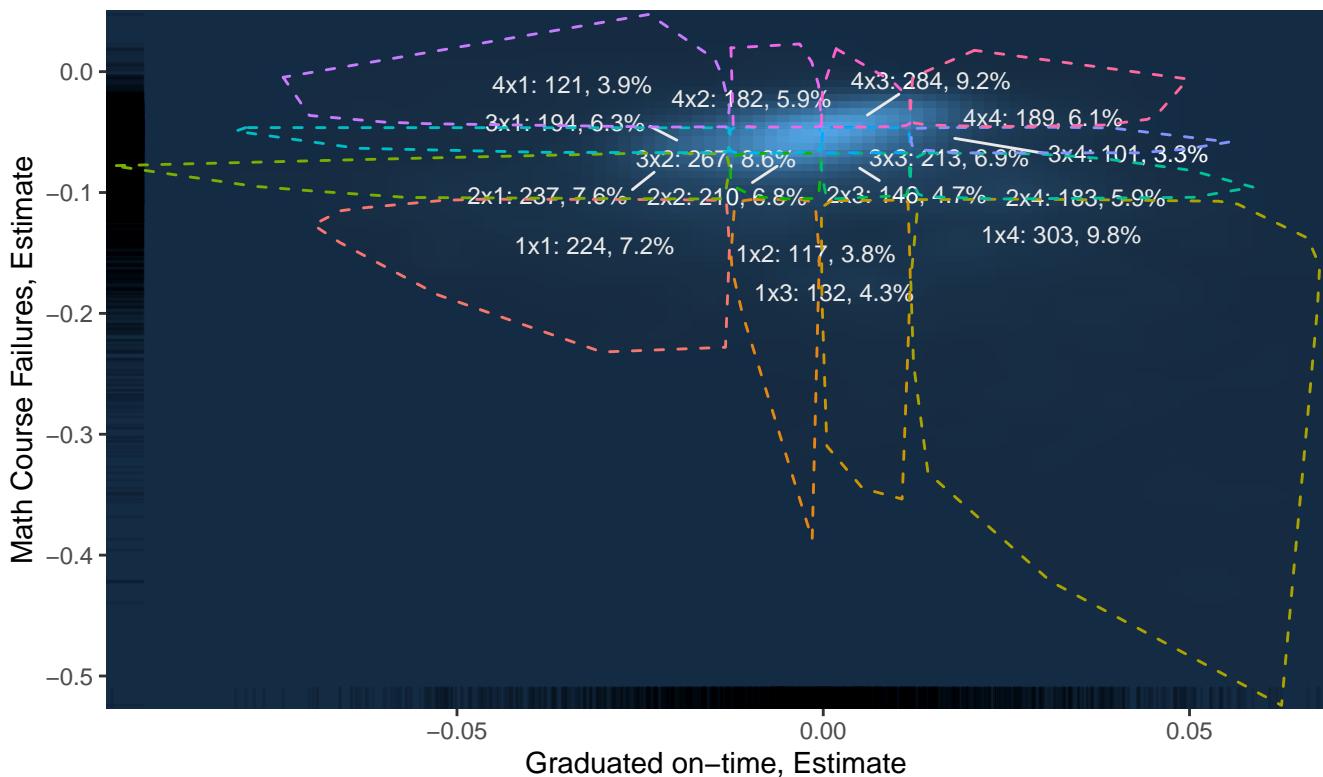
PTE Decile Plot: Math Course Failures vs. Graduated on-time



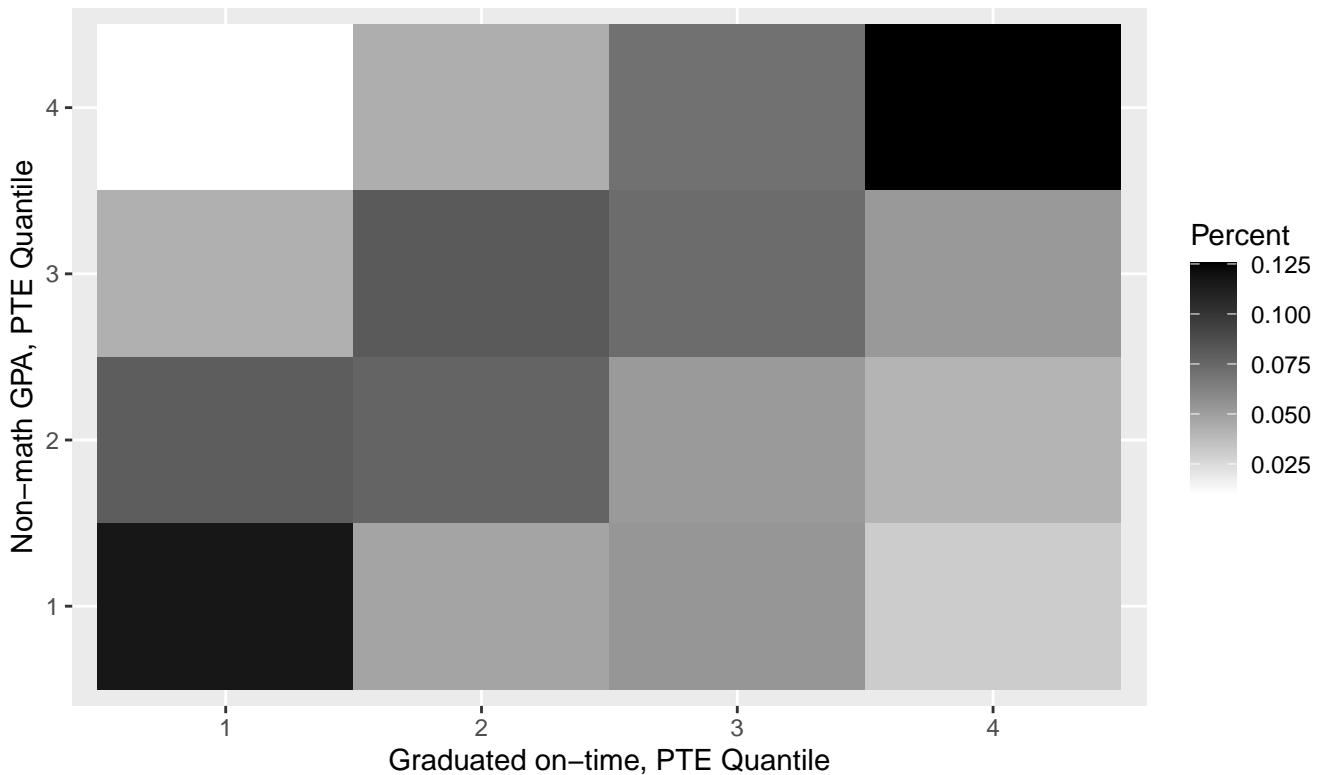
PTE Scatterplot: Math Course Failures vs Graduated on-time



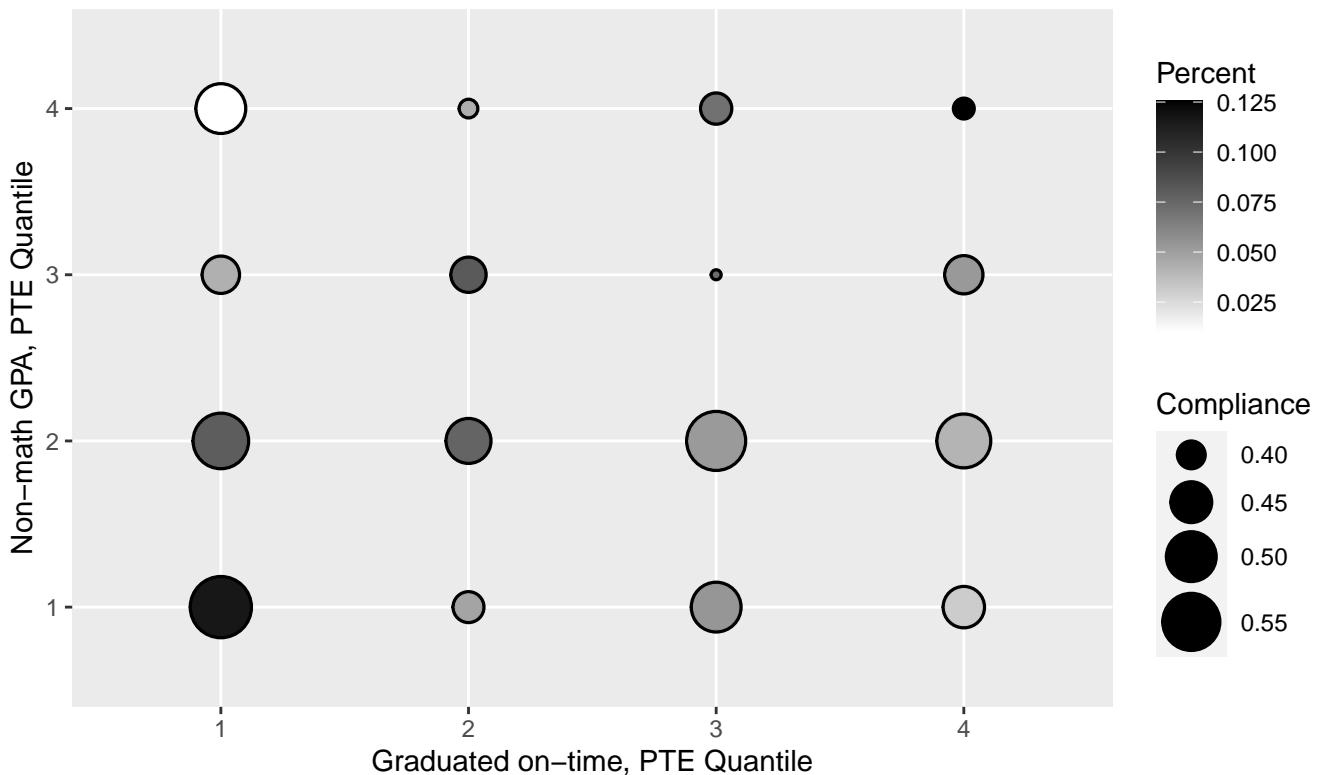
PTE Densities: Math Course Failures vs Graduated on-time



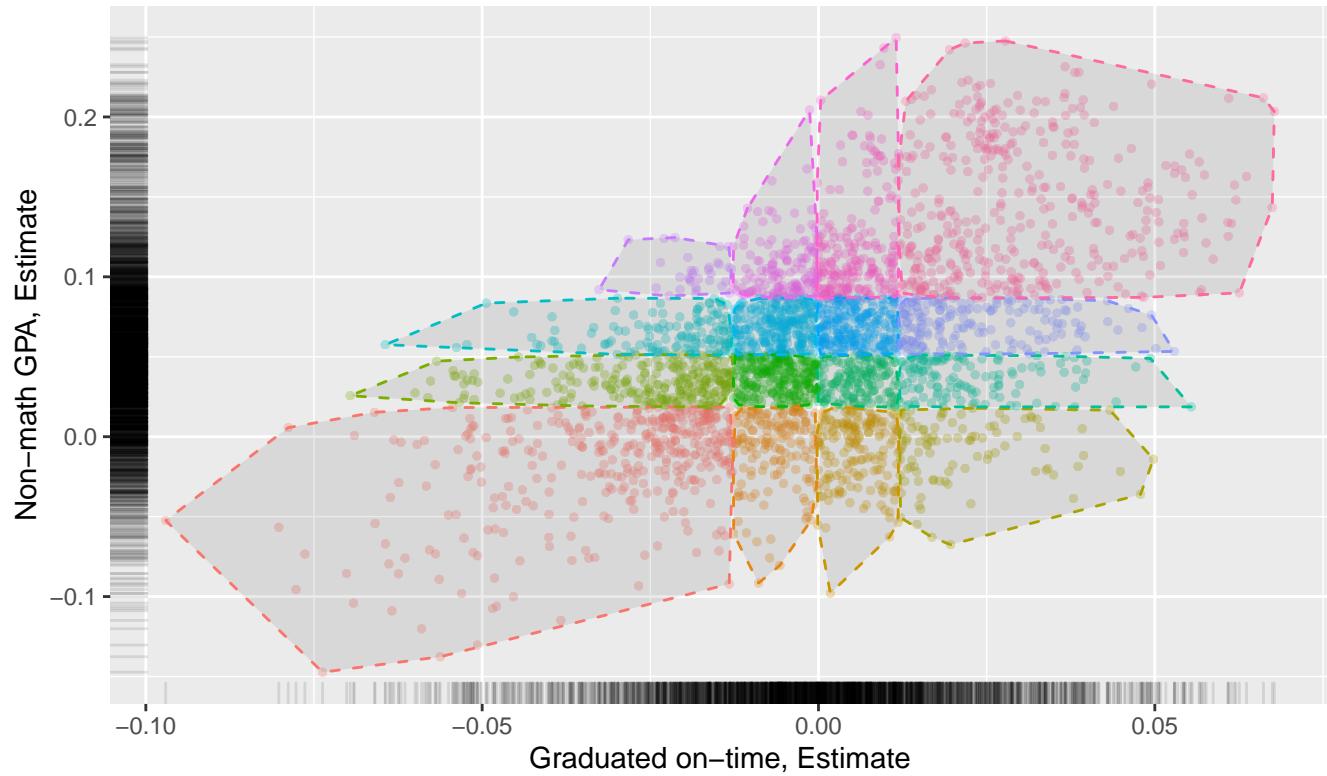
PTE Decile Plot: Non–math GPA vs. Graduated on–time



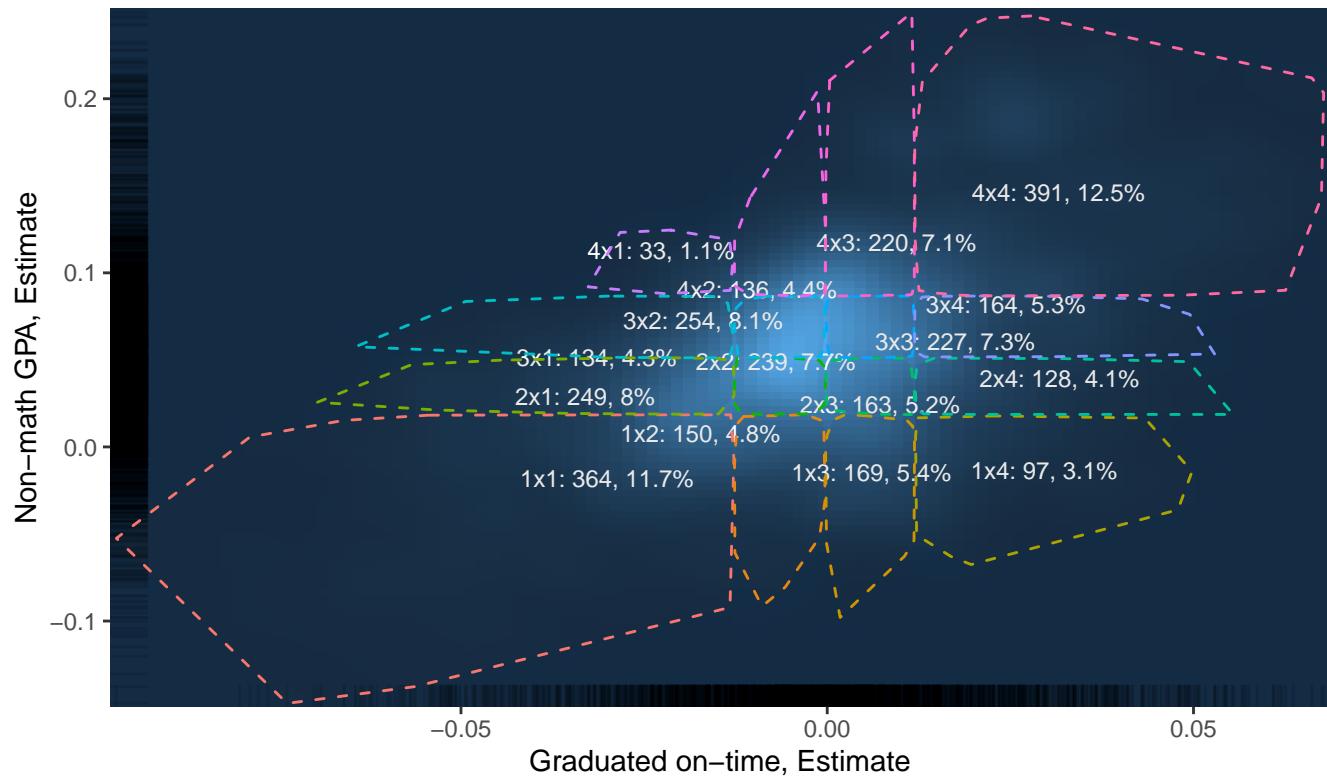
PTE Decile Plot: Non–math GPA vs. Graduated on–time



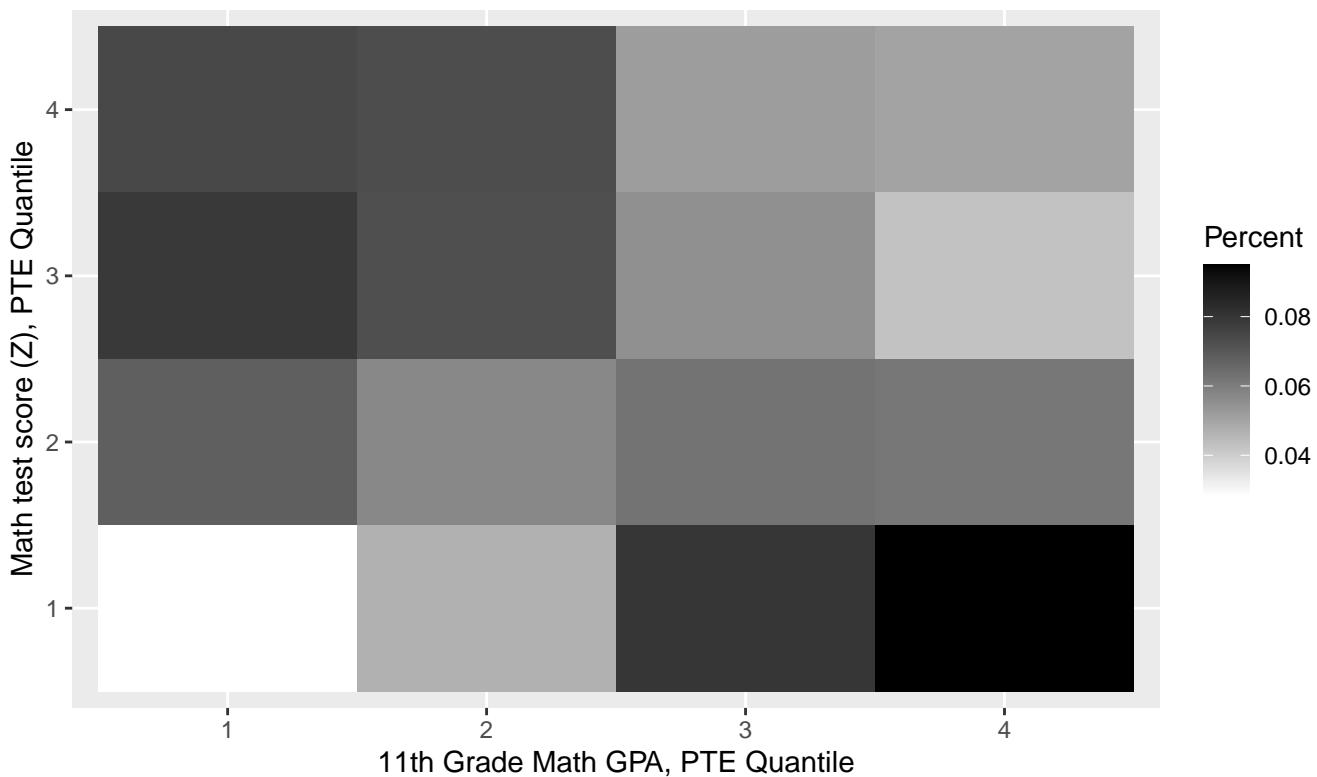
PTE Scatterplot: Non–math GPA vs Graduated on–time



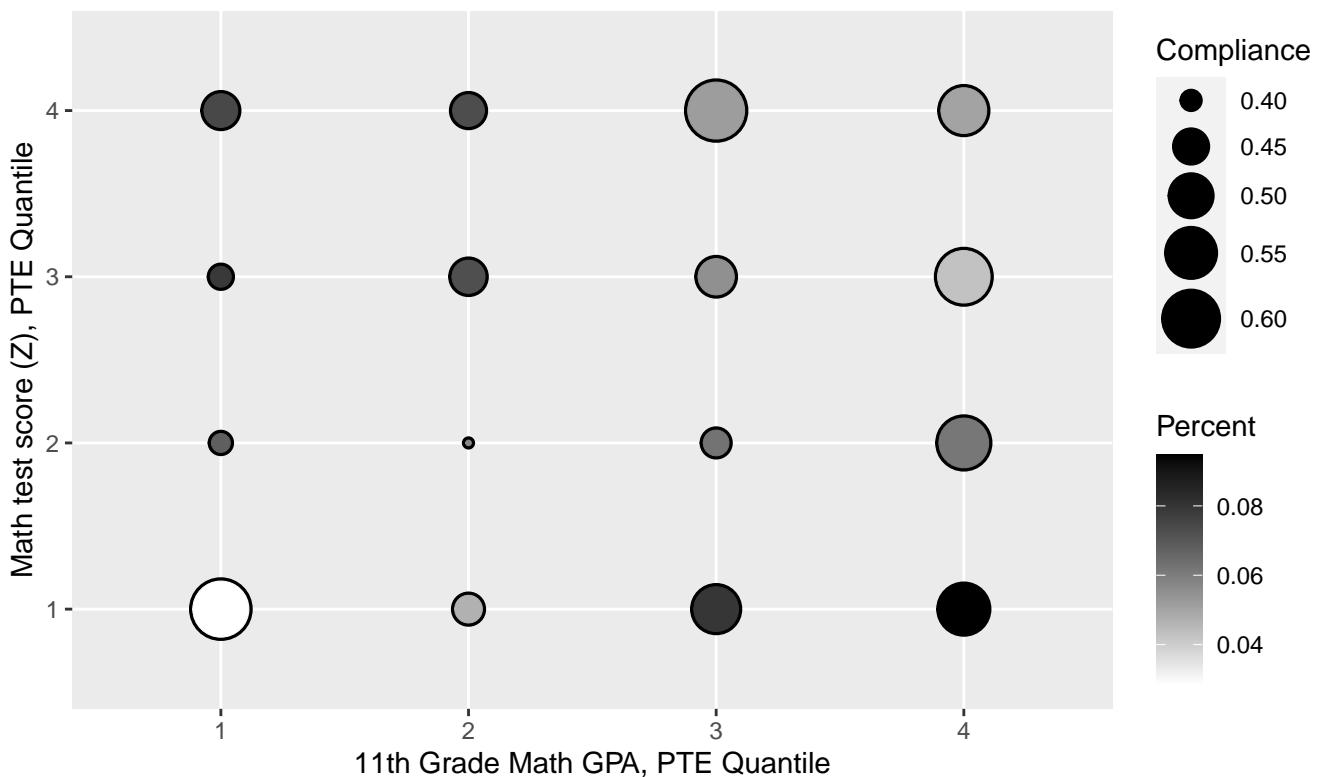
PTE Densities: Non–math GPA vs Graduated on–time



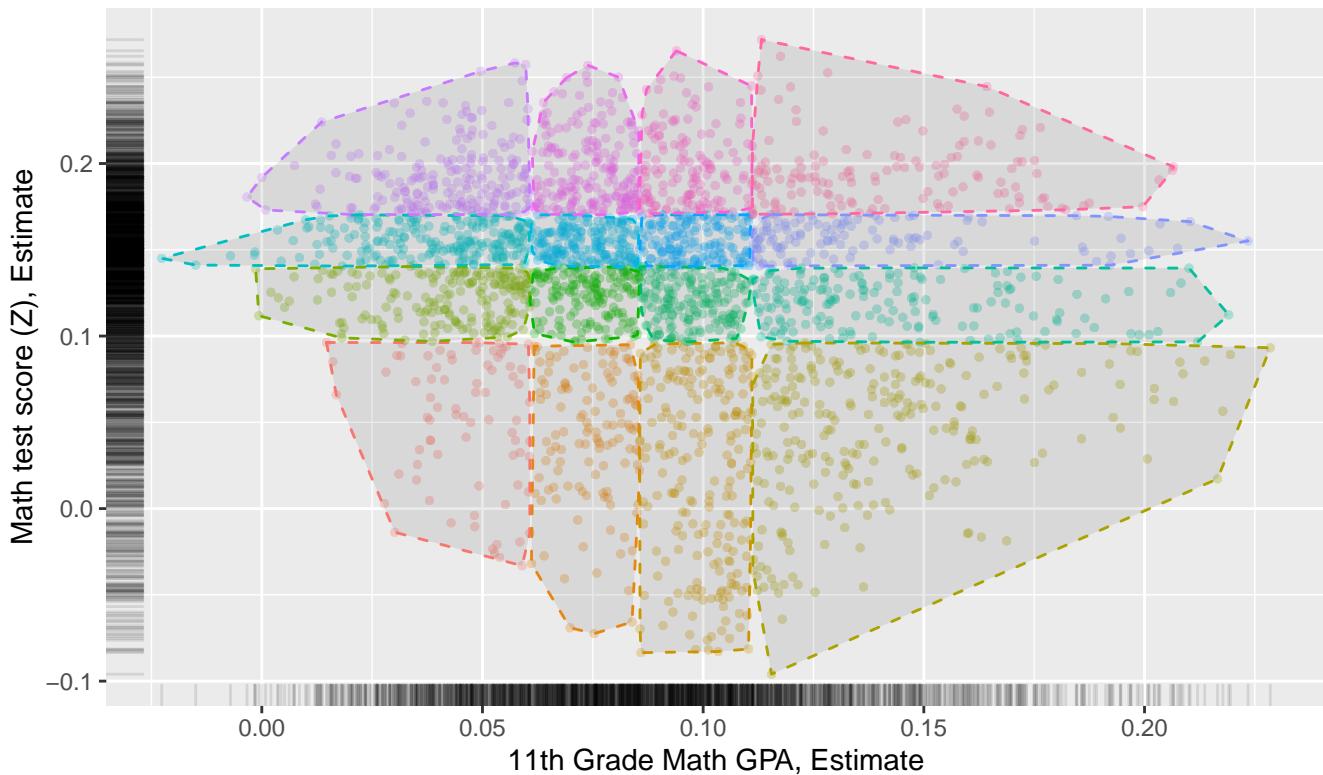
PTE Decile Plot: Math test score (Z) vs. 11th Grade Math GPA



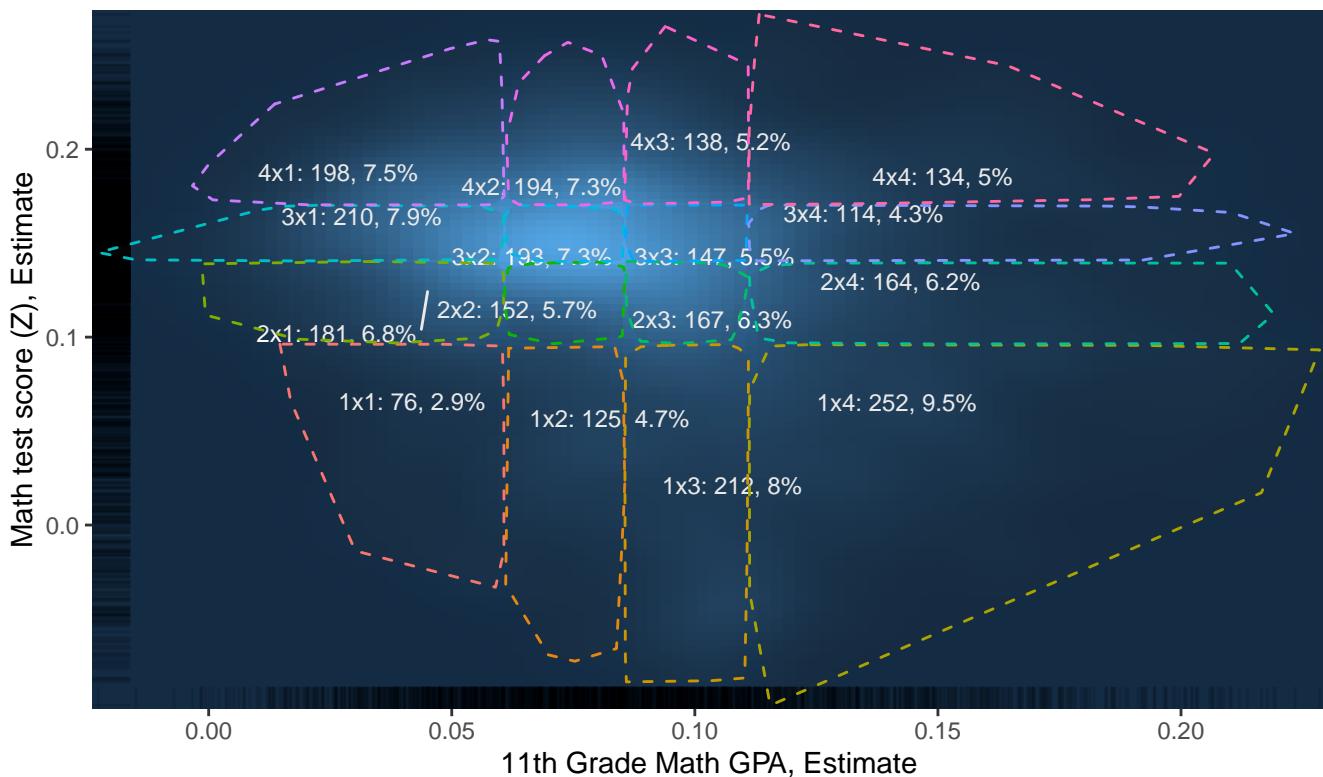
PTE Decile Plot: Math test score (Z) vs. 11th Grade Math GPA



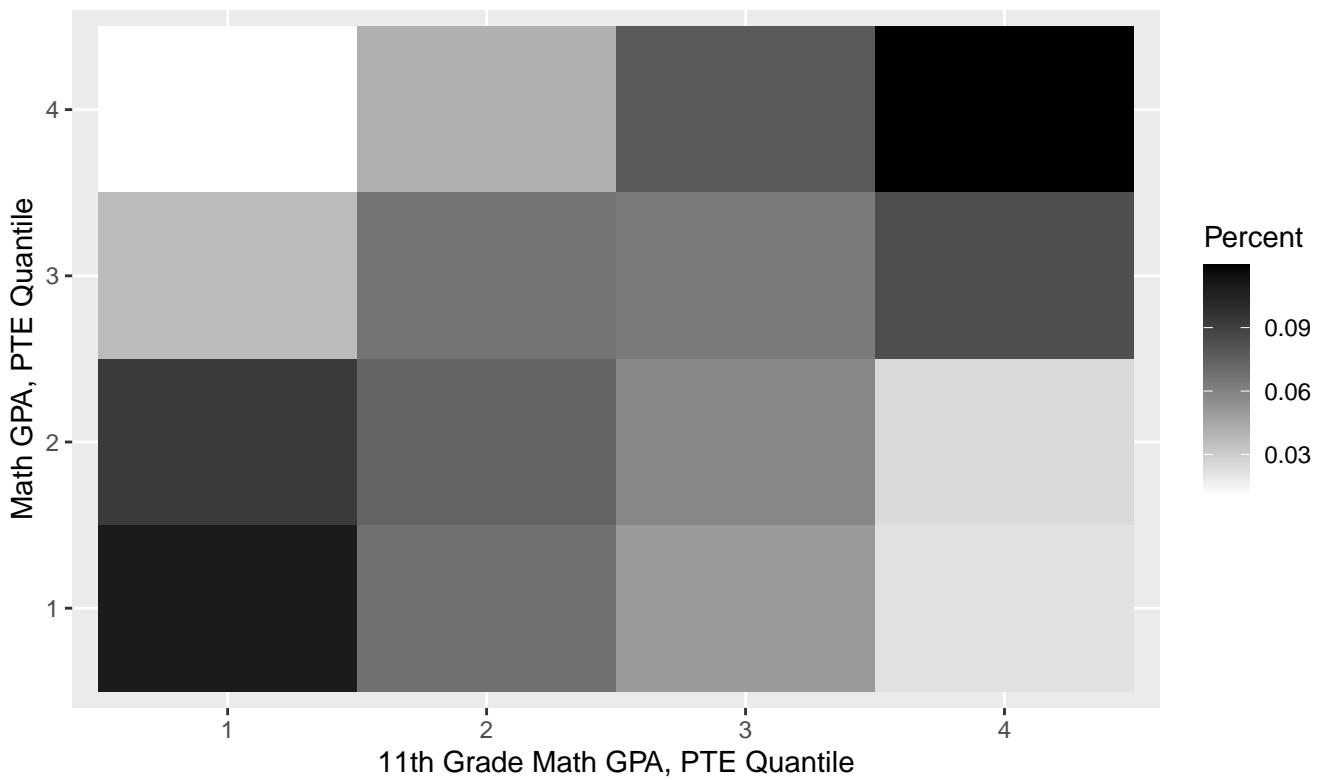
PTE Scatterplot: Math test score (Z) vs 11th Grade Math GPA



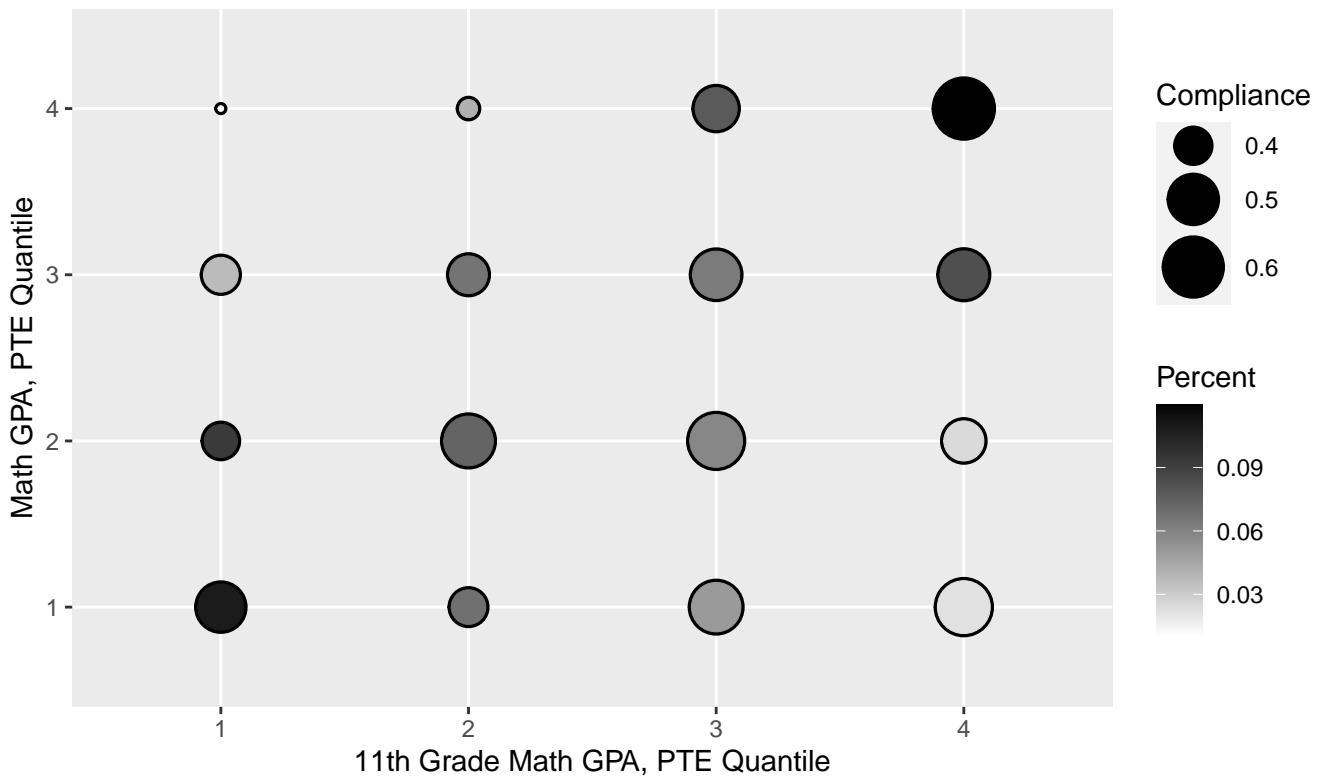
PTE Densities: Math test score (Z) vs 11th Grade Math GPA



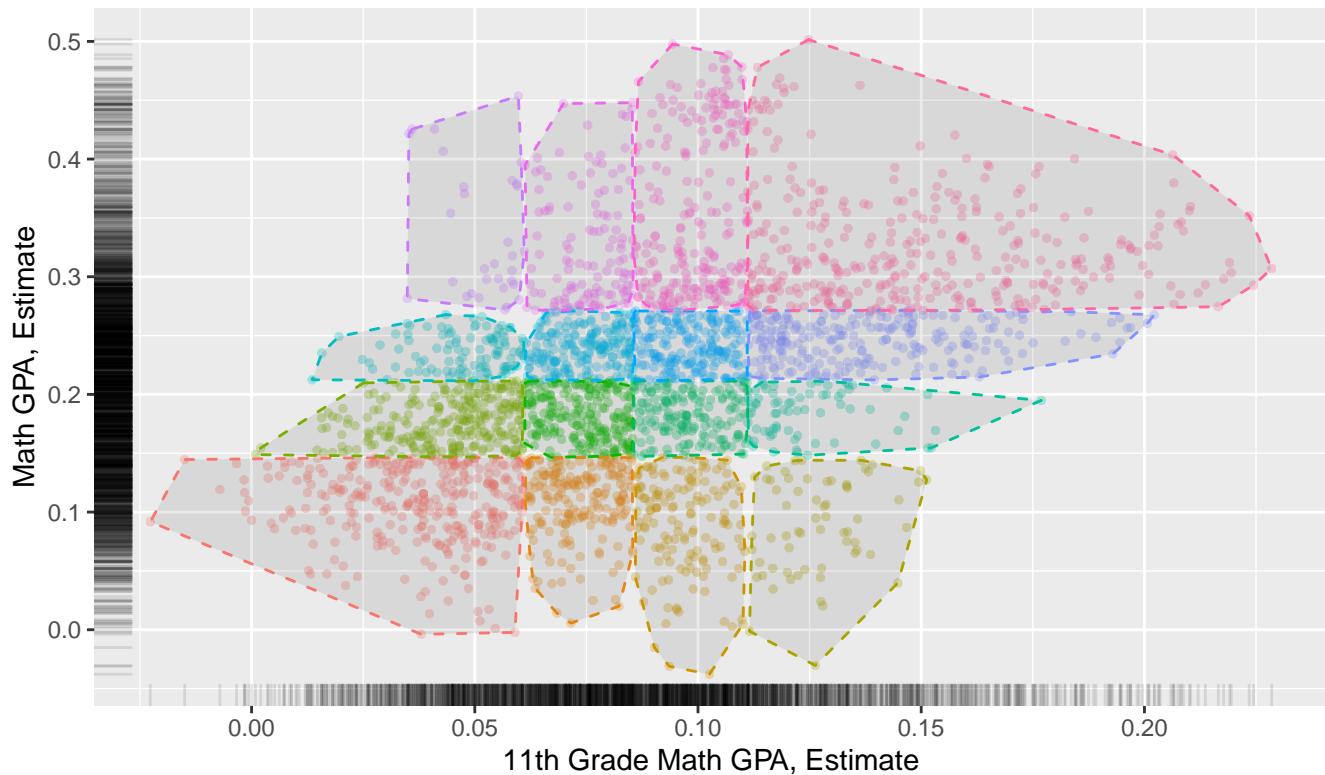
PTE Decile Plot: Math GPA vs. 11th Grade Math GPA



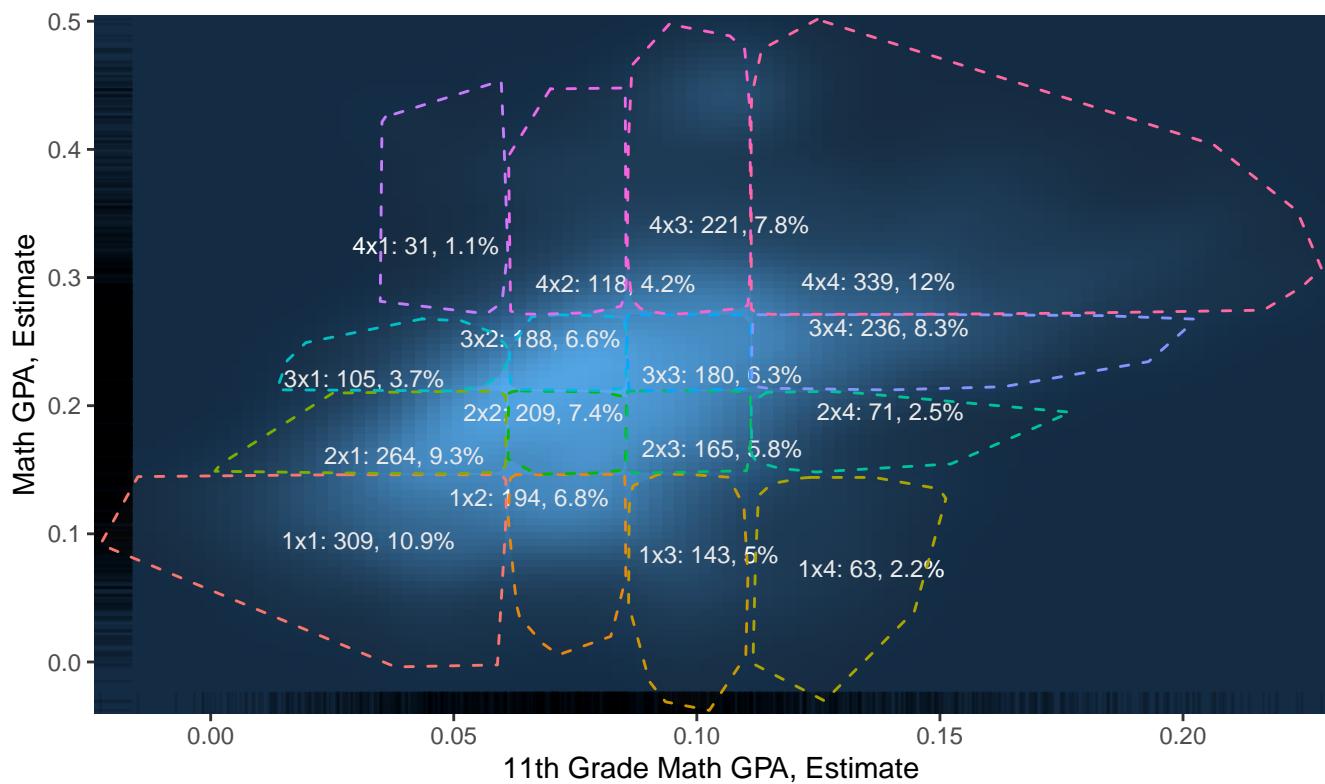
PTE Decile Plot: Math GPA vs. 11th Grade Math GPA



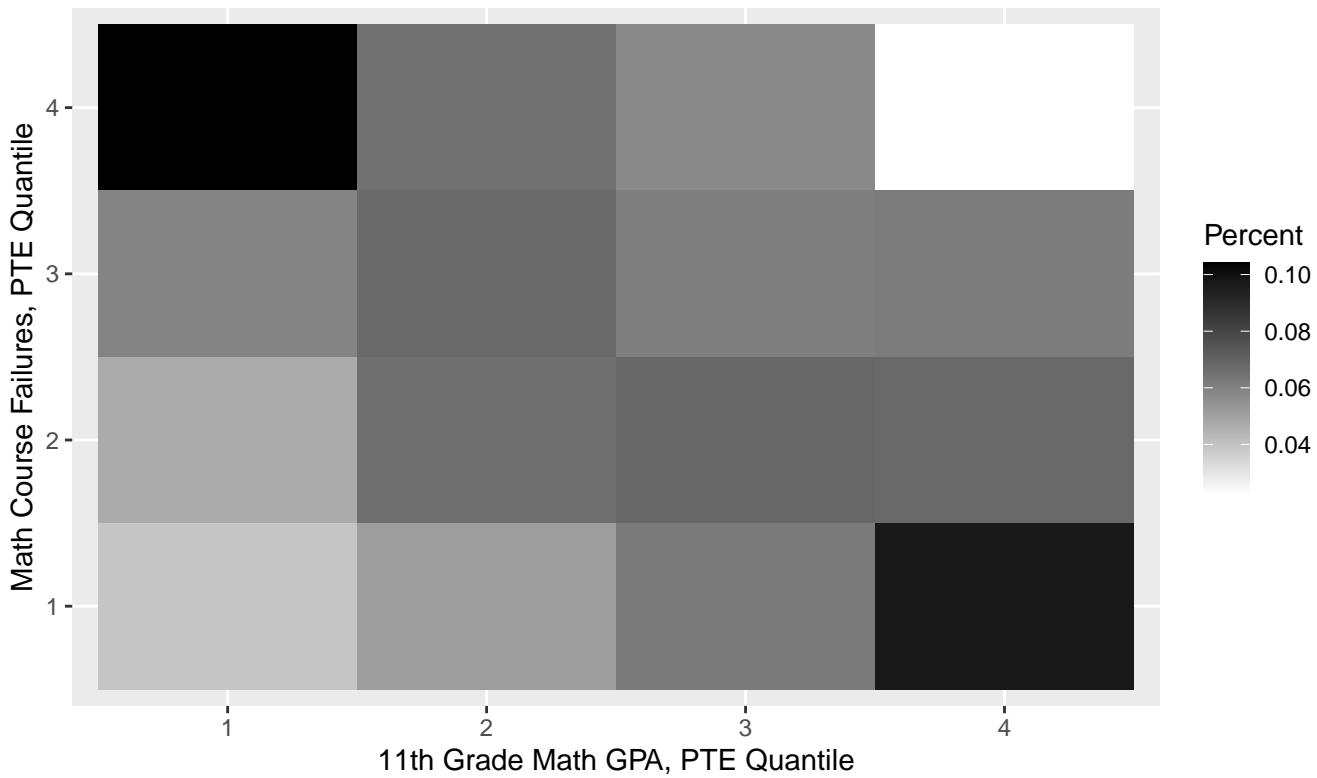
PTE Scatterplot: Math GPA vs 11th Grade Math GPA



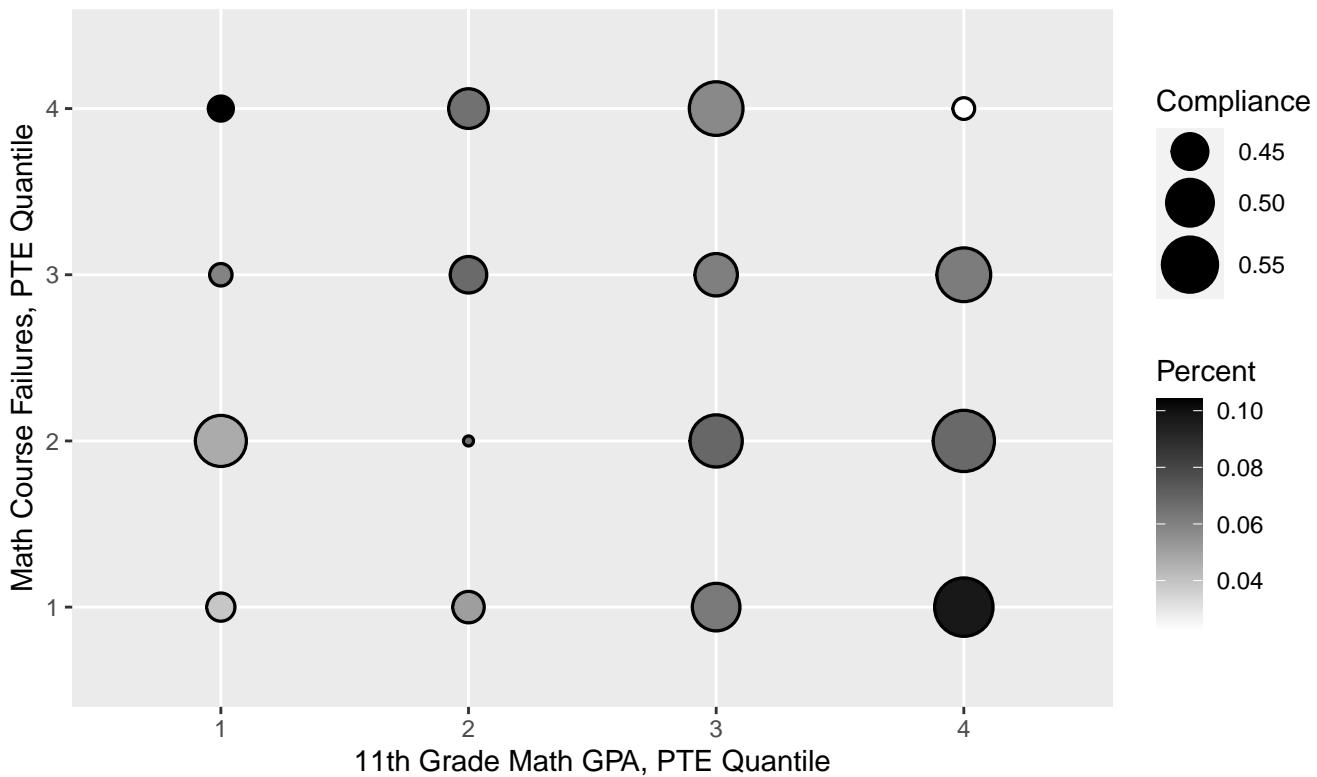
PTE Densities: Math GPA vs 11th Grade Math GPA



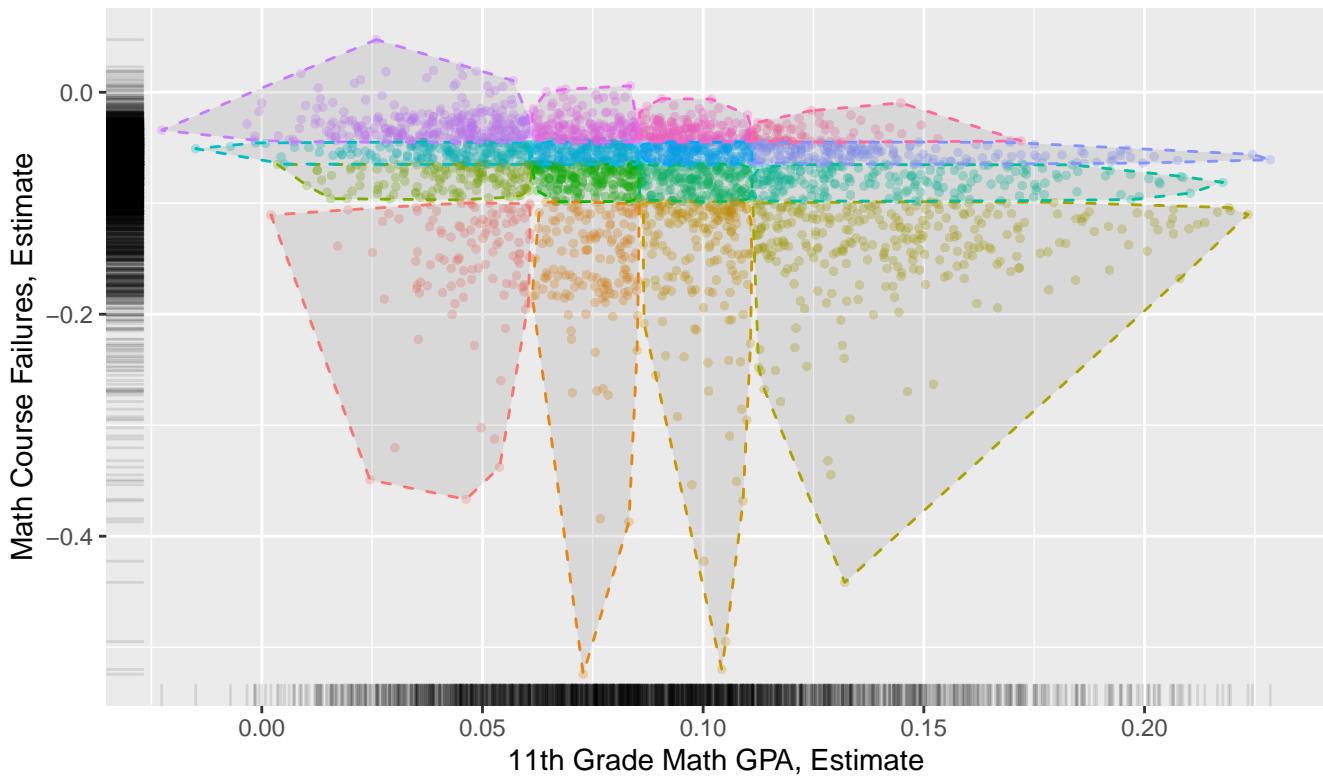
PTE Decile Plot: Math Course Failures vs. 11th Grade Math GPA



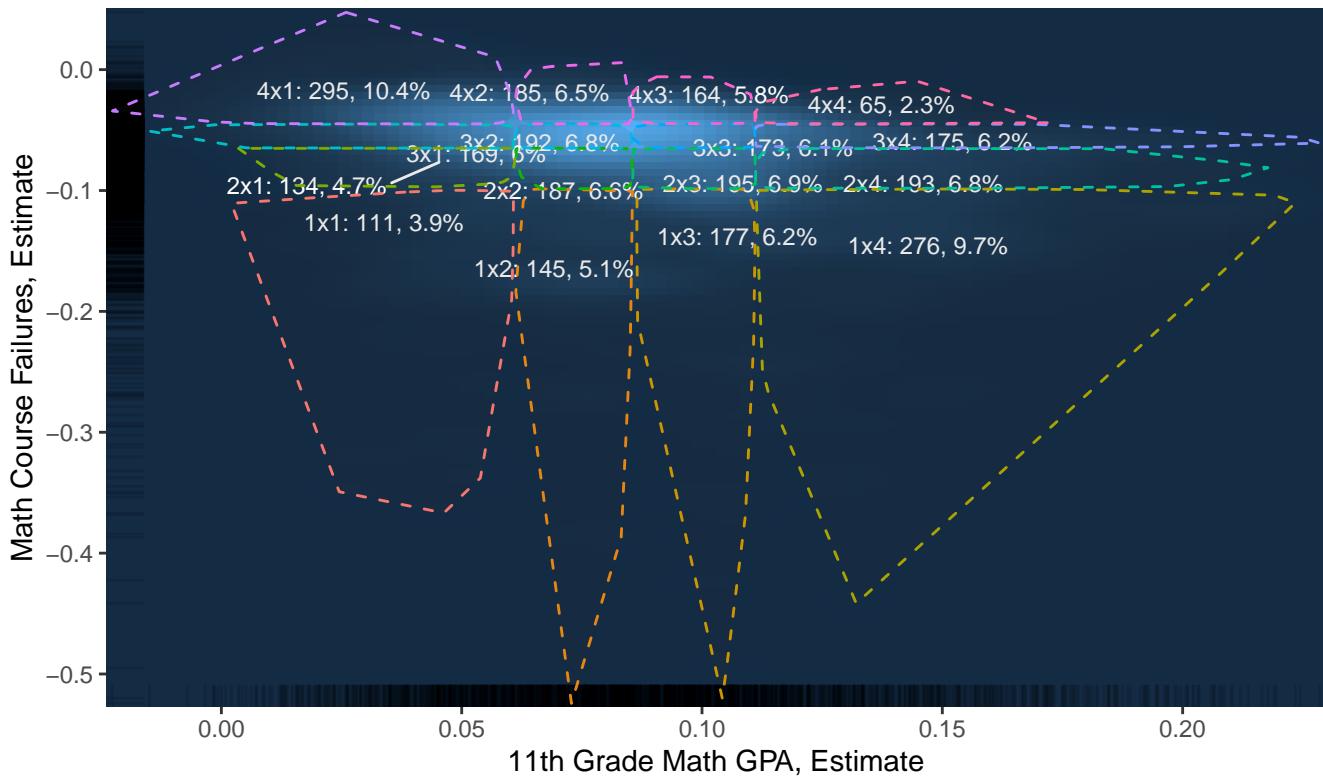
PTE Decile Plot: Math Course Failures vs. 11th Grade Math GPA



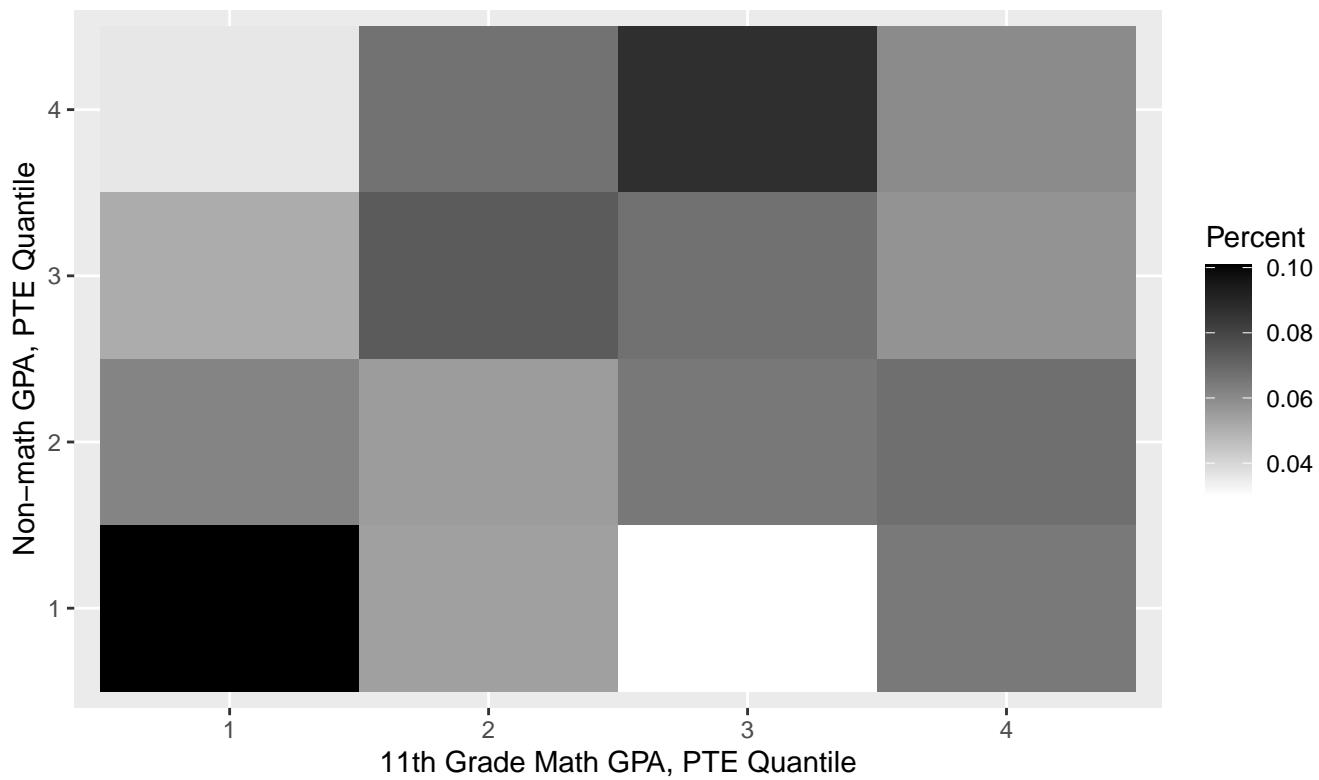
PTE Scatterplot: Math Course Failures vs 11th Grade Math GPA



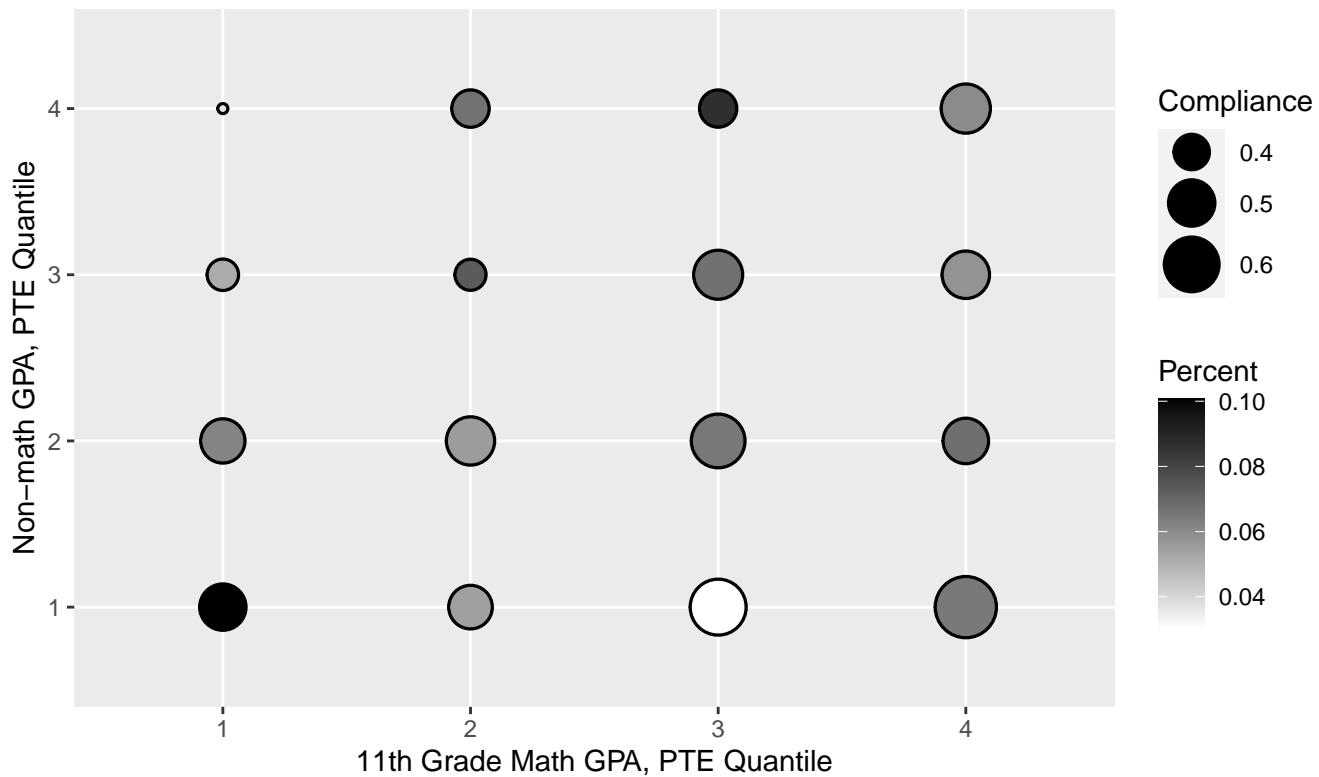
PTE Densities: Math Course Failures vs 11th Grade Math GPA



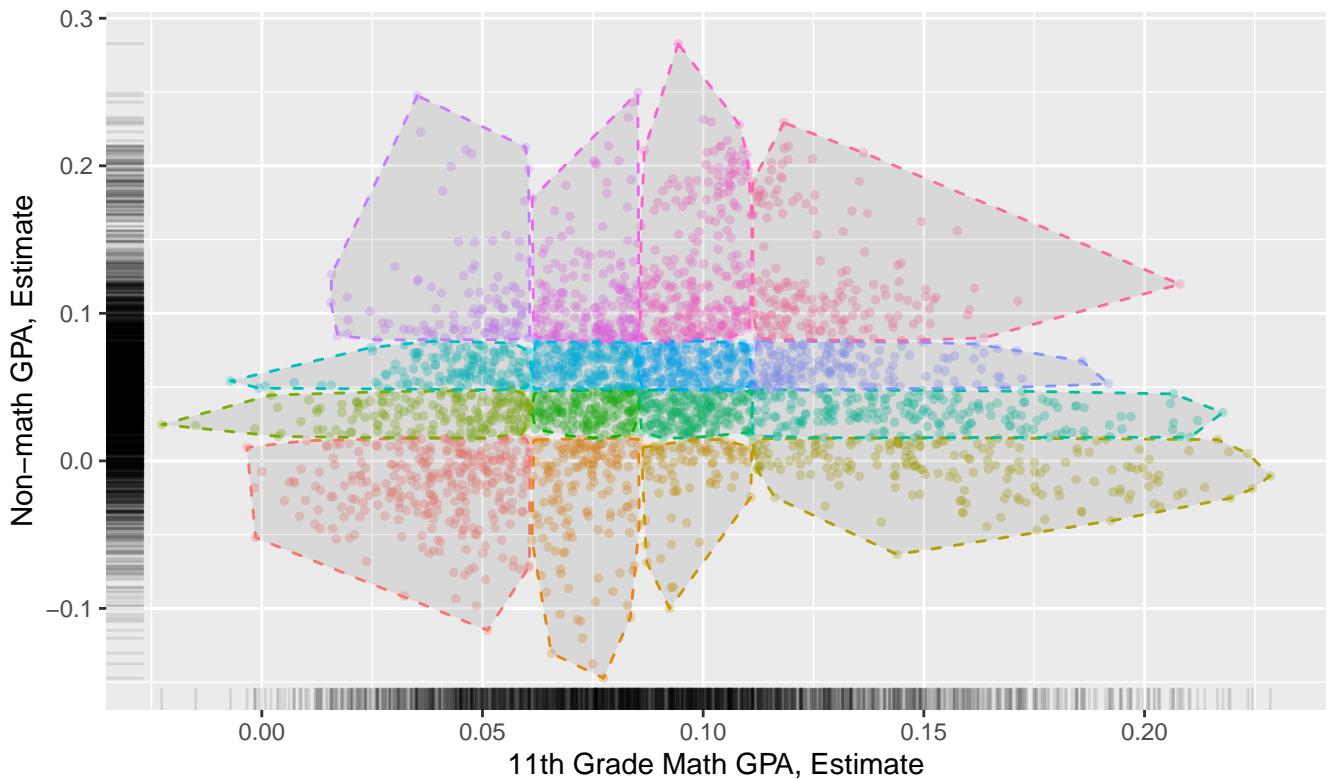
PTE Decile Plot: Non–math GPA vs. 11th Grade Math GPA



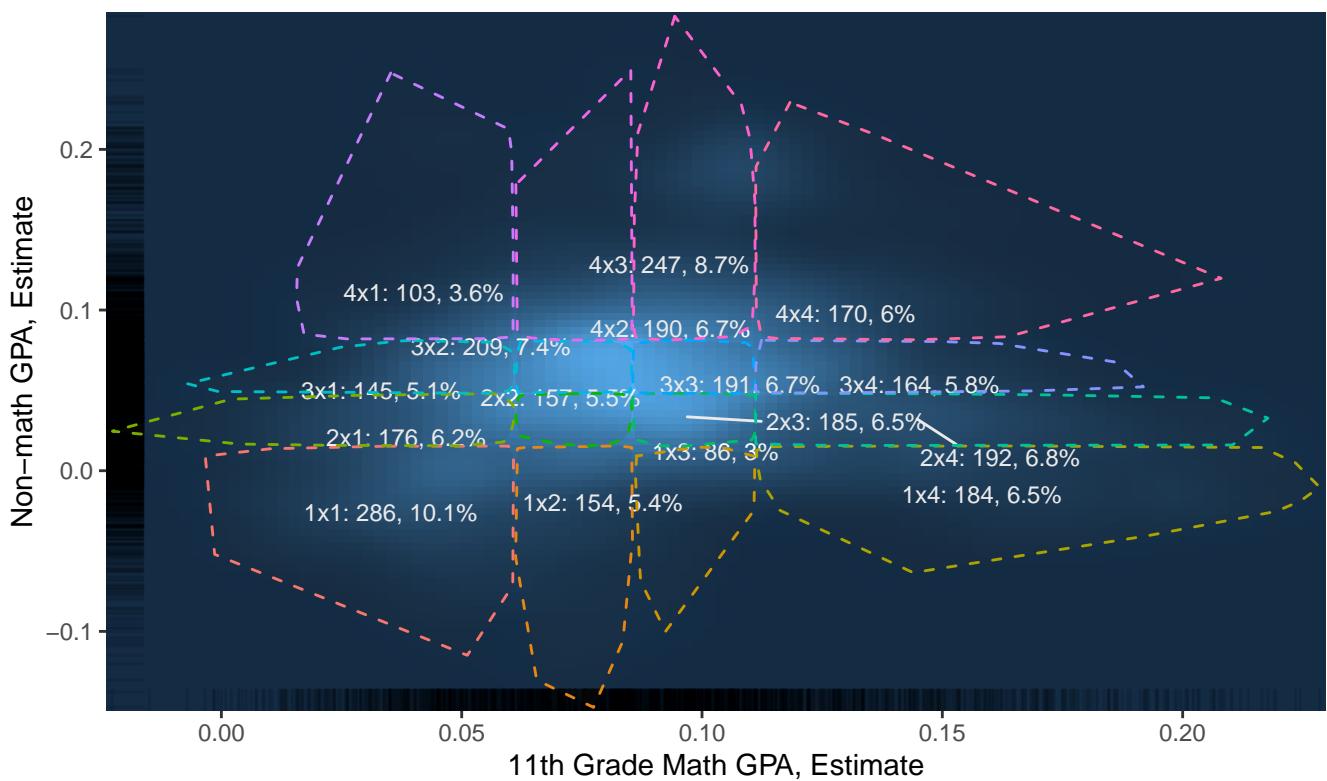
PTE Decile Plot: Non–math GPA vs. 11th Grade Math GPA



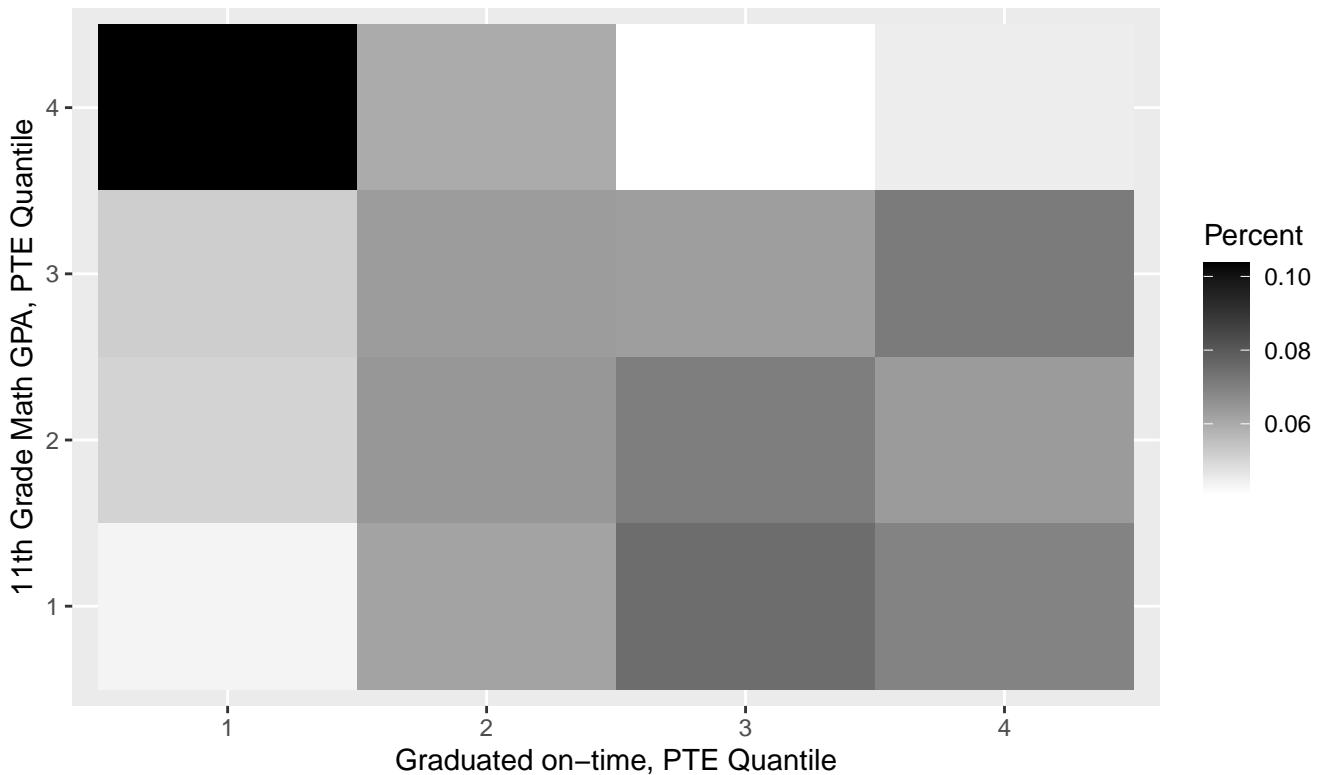
PTE Scatterplot: Non–math GPA vs 11th Grade Math GPA



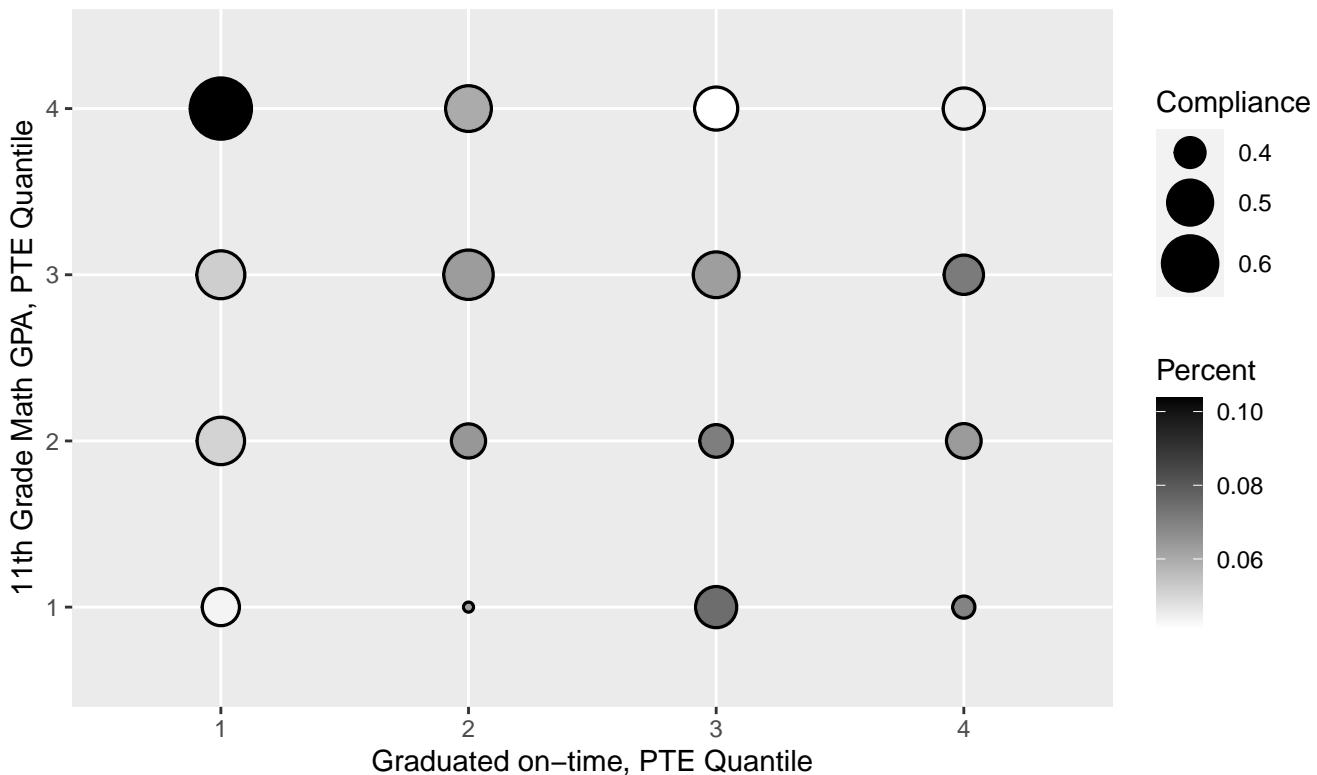
PTE Densities: Non–math GPA vs 11th Grade Math GPA



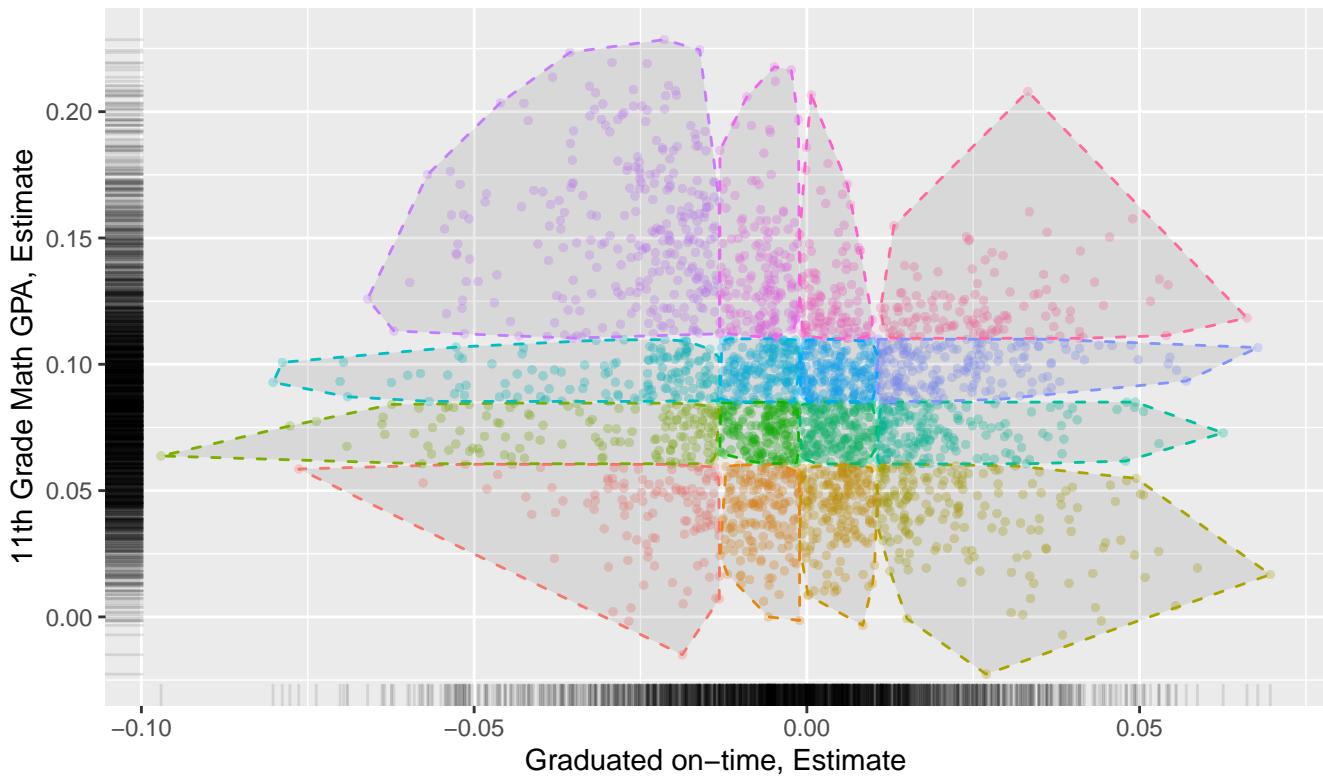
PTE Decile Plot: 11th Grade Math GPA vs. Graduated on-time



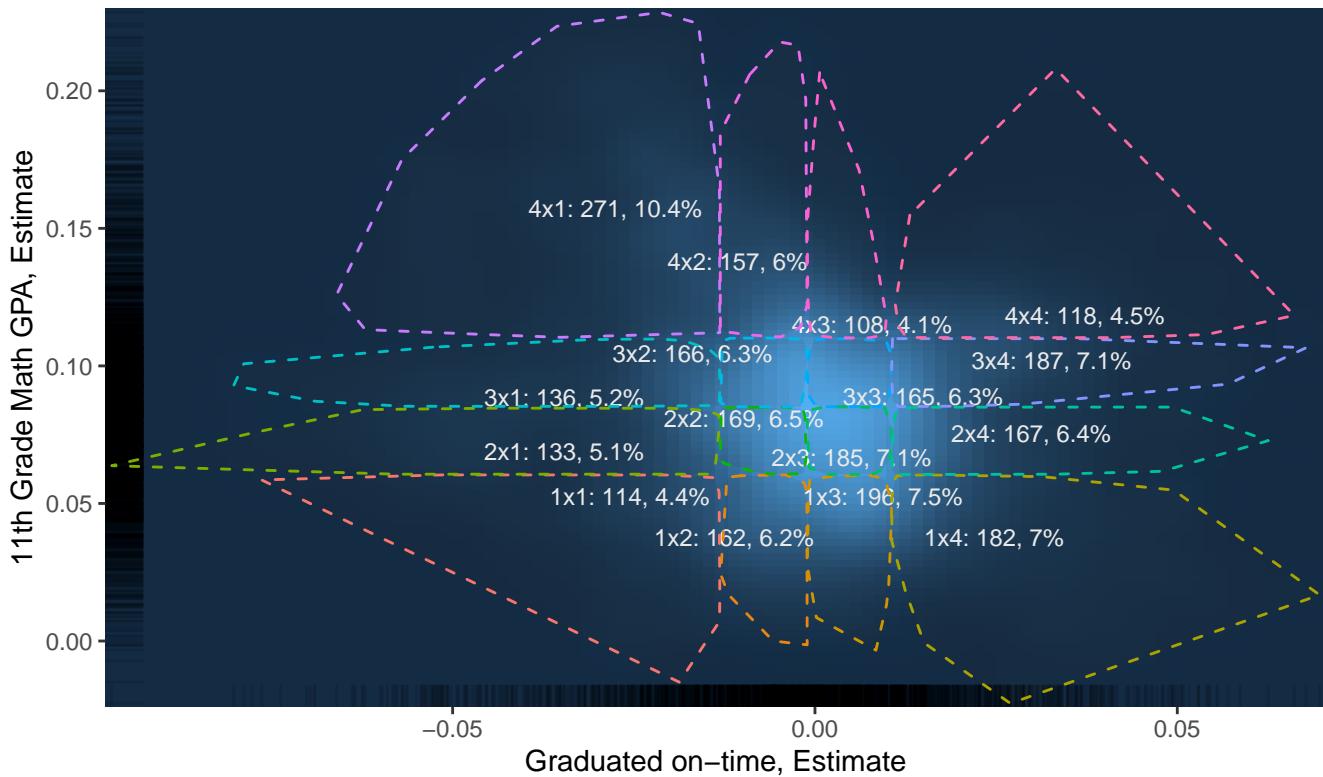
PTE Decile Plot: 11th Grade Math GPA vs. Graduated on-time



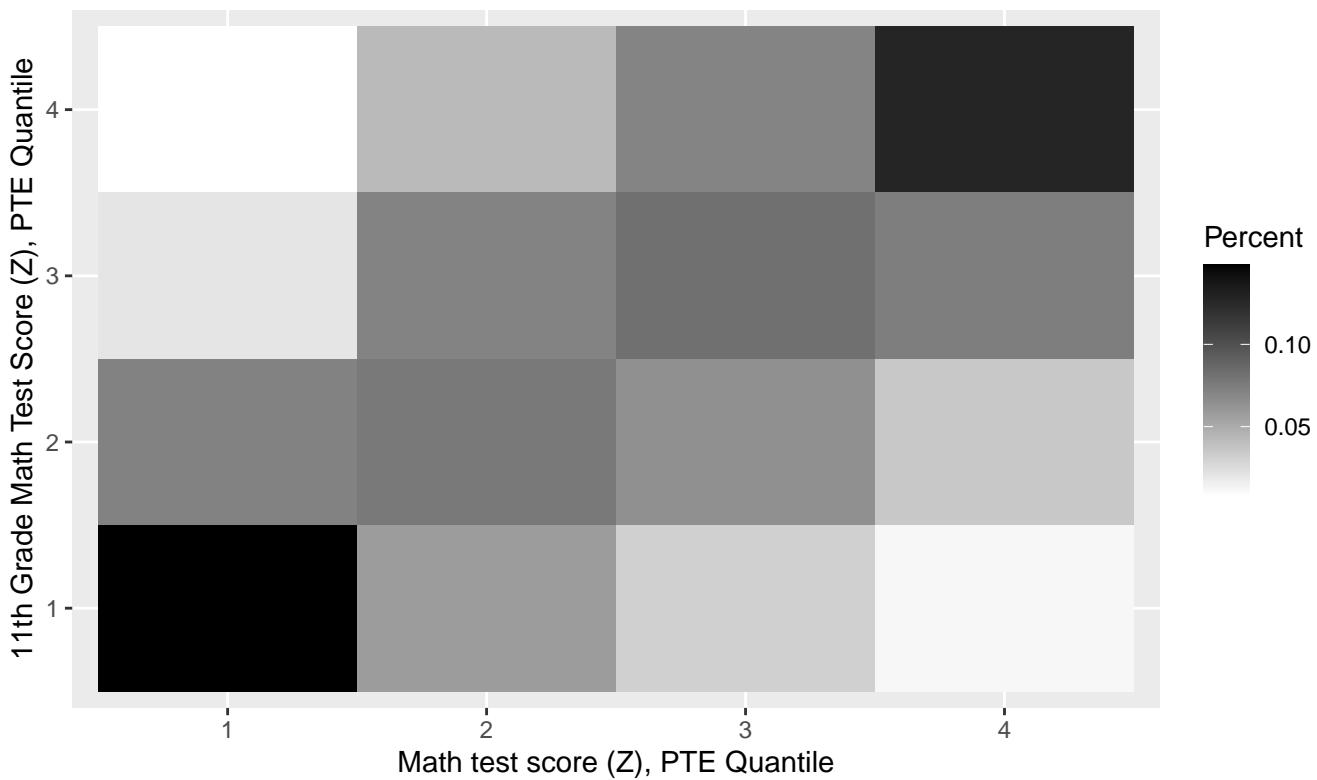
PTE Scatterplot: 11th Grade Math GPA vs Graduated on-time



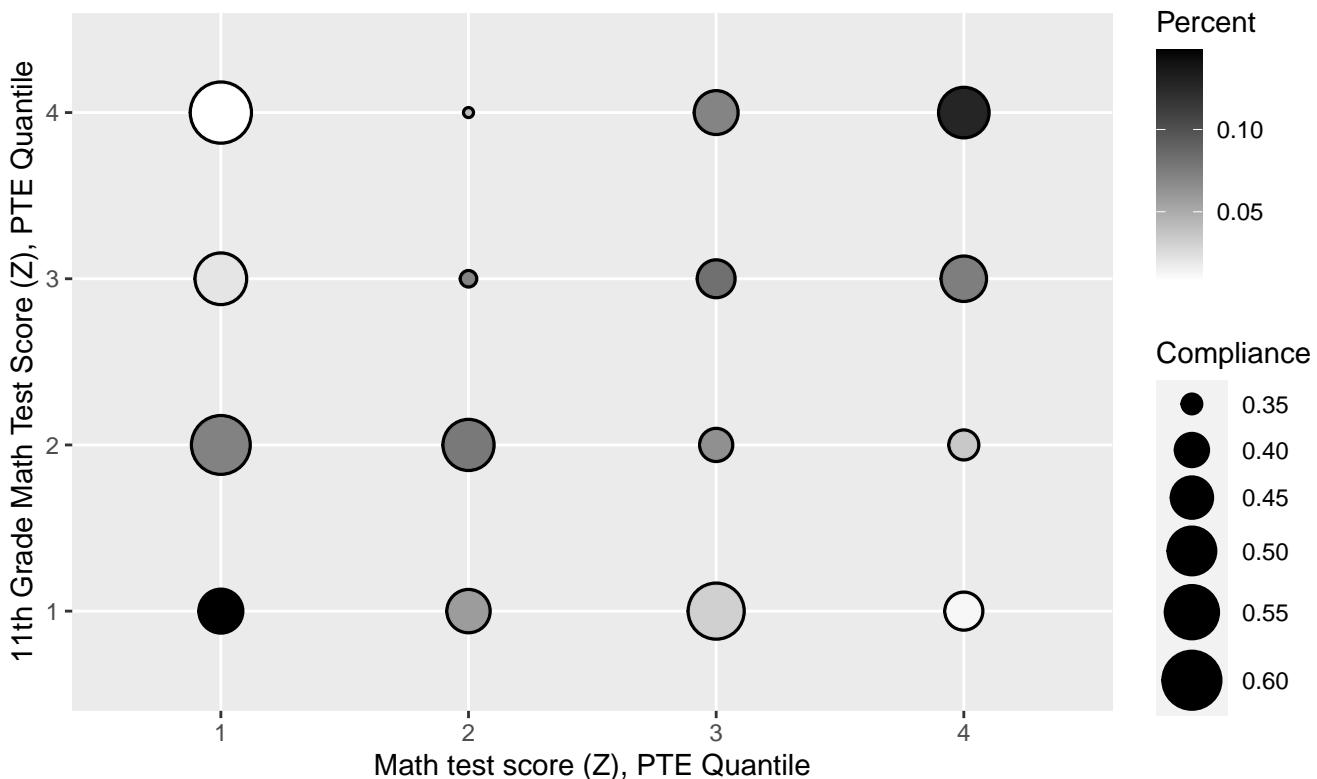
PTE Densities: 11th Grade Math GPA vs Graduated on-time



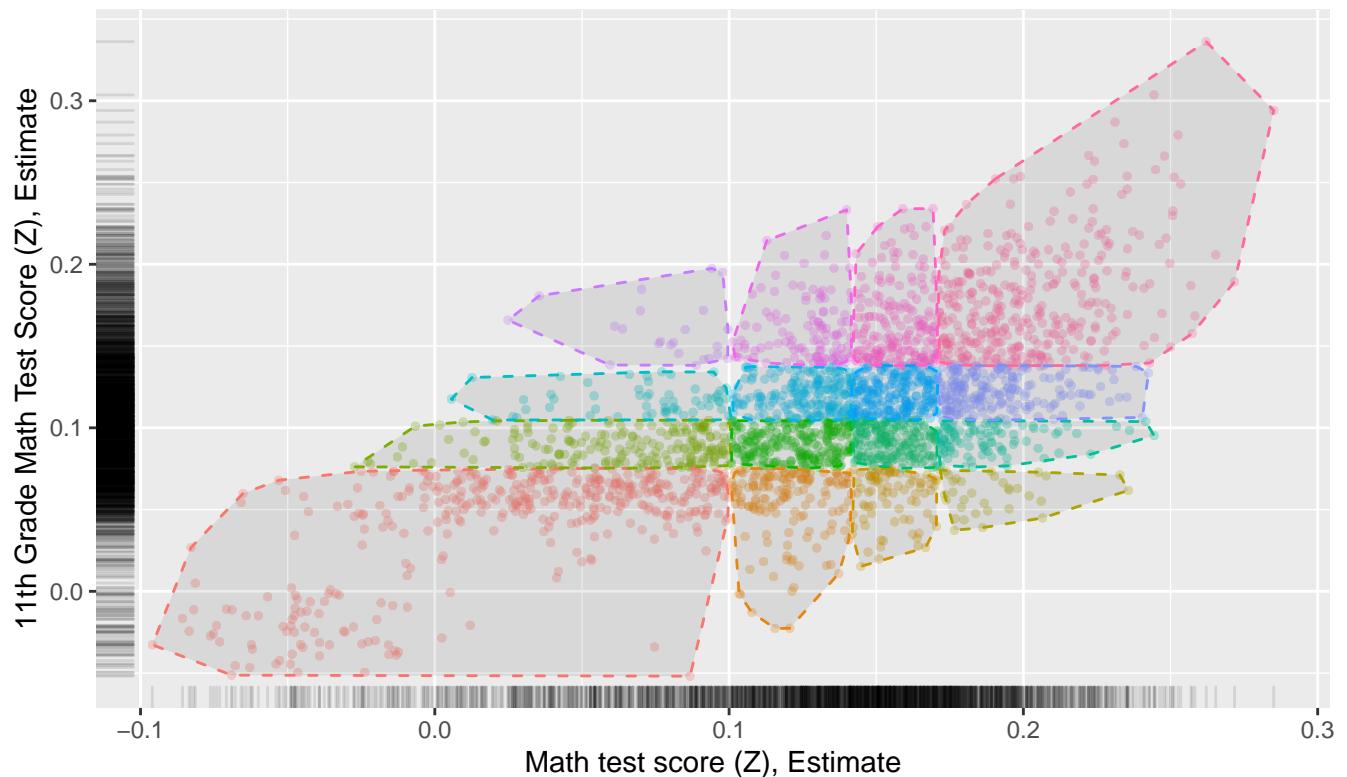
PTE Decile Plot: 11th Grade Math Test Score ( $Z$ ) vs. Math test score ( $Z$ )



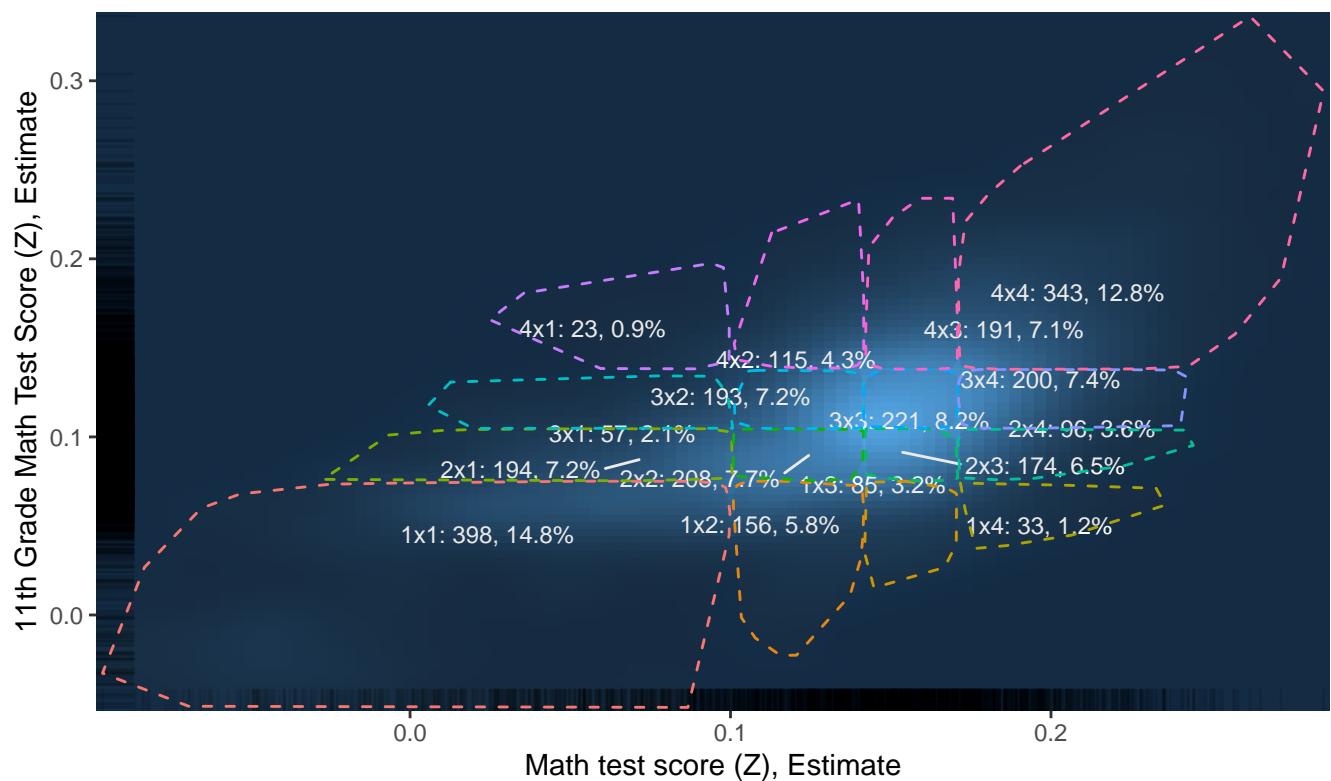
PTE Decile Plot: 11th Grade Math Test Score ( $Z$ ) vs. Math test score ( $Z$ )



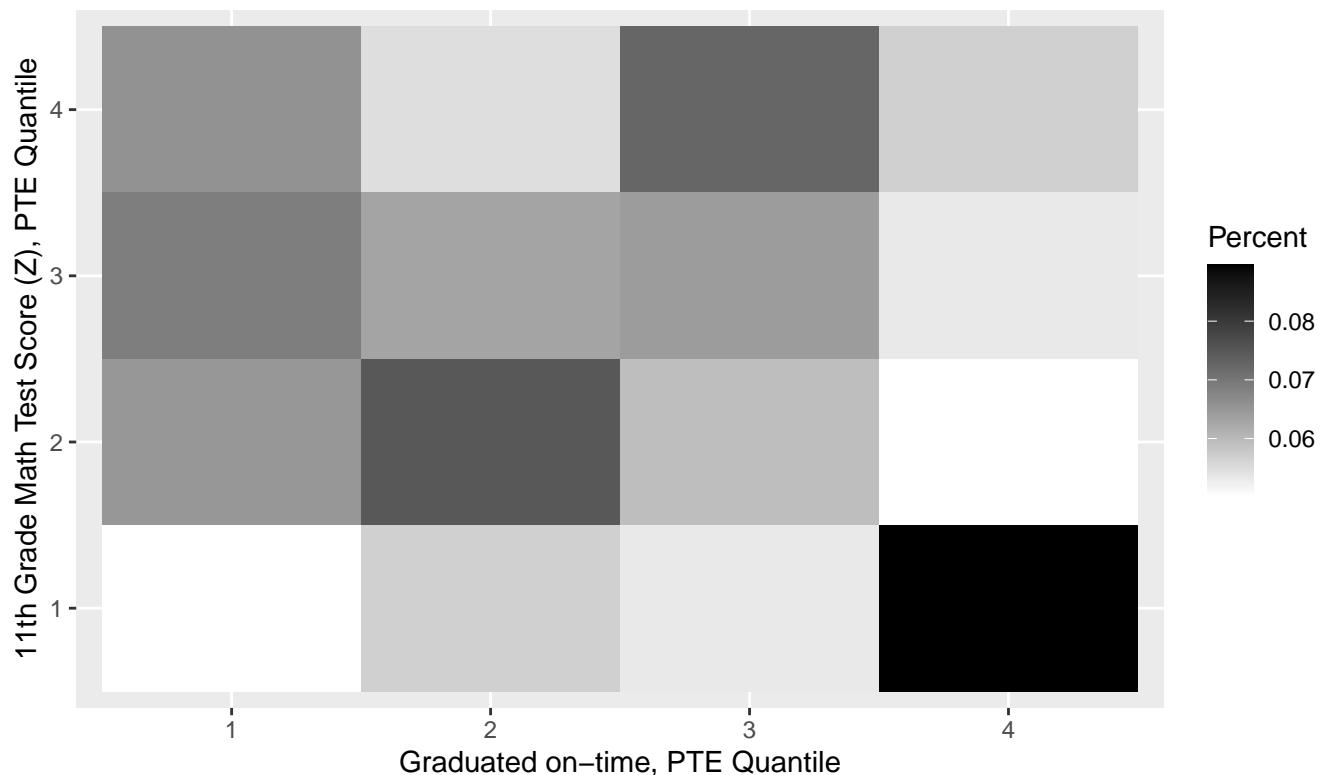
PTE Scatterplot: 11th Grade Math Test Score (Z) vs Math test score (Z)



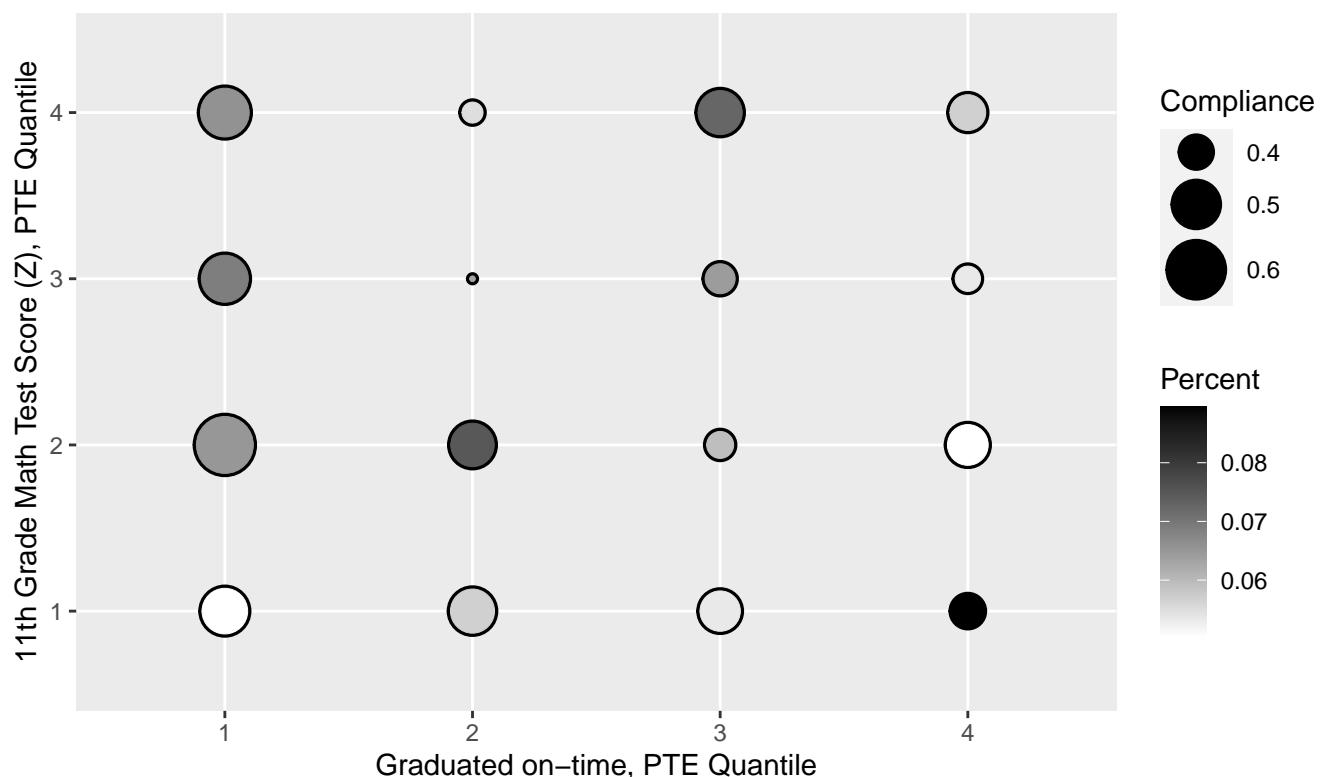
PTE Densities: 11th Grade Math Test Score (Z) vs Math test score (Z)



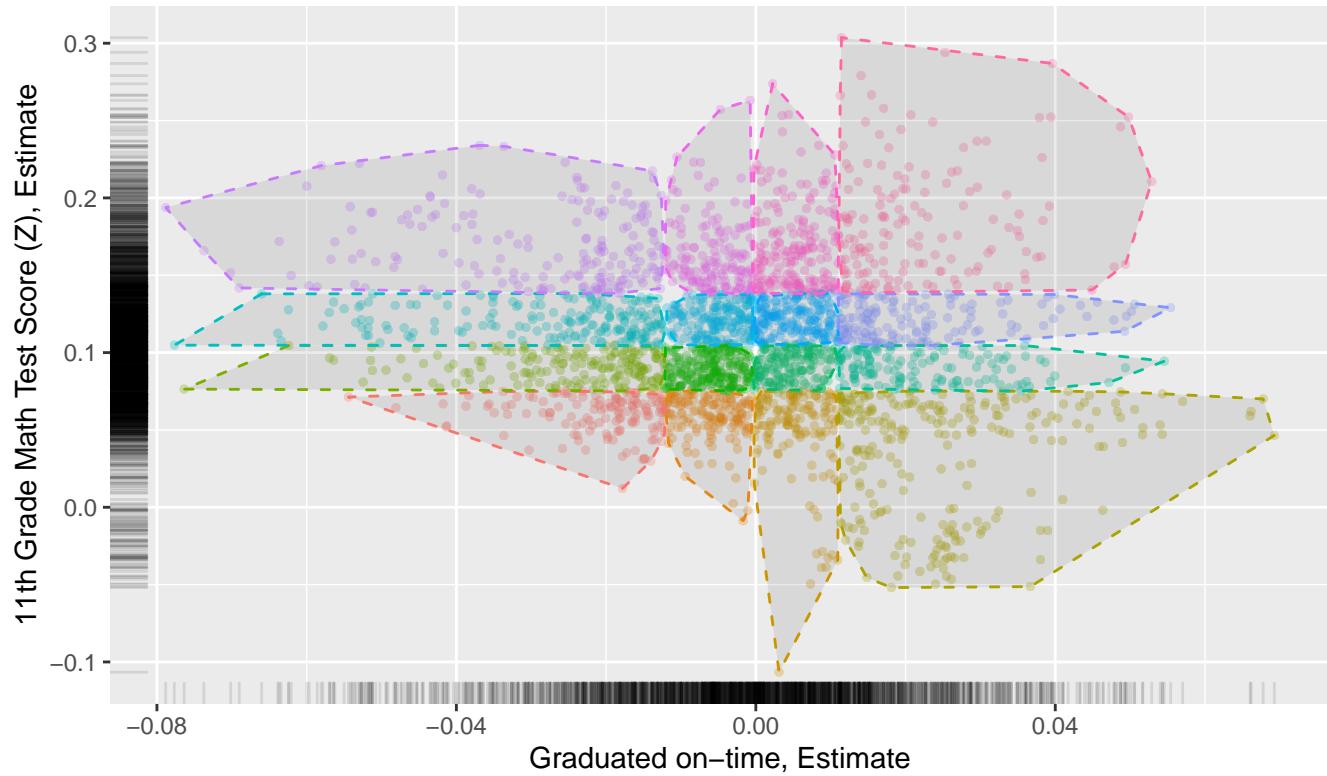
PTE Decile Plot: 11th Grade Math Test Score ( $Z$ ) vs. Graduated on-time



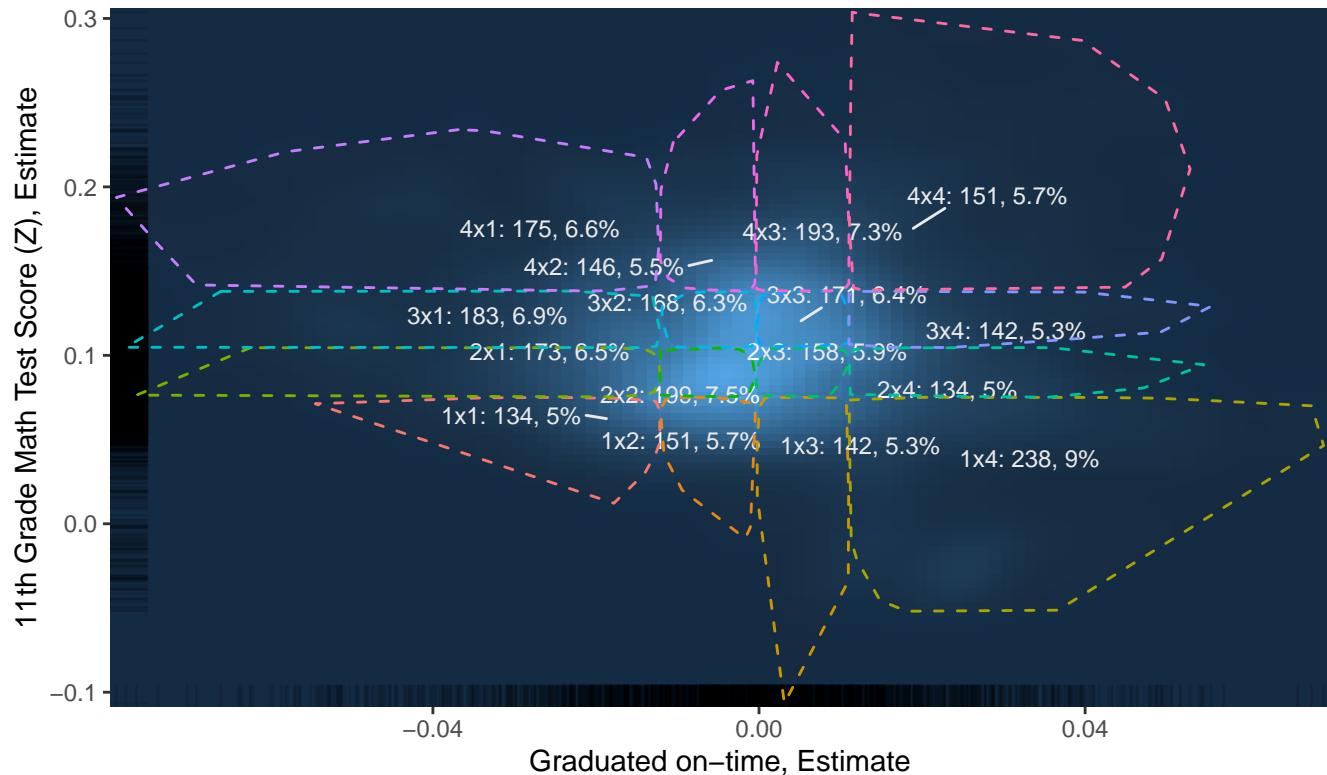
PTE Decile Plot: 11th Grade Math Test Score ( $Z$ ) vs. Graduated on-time



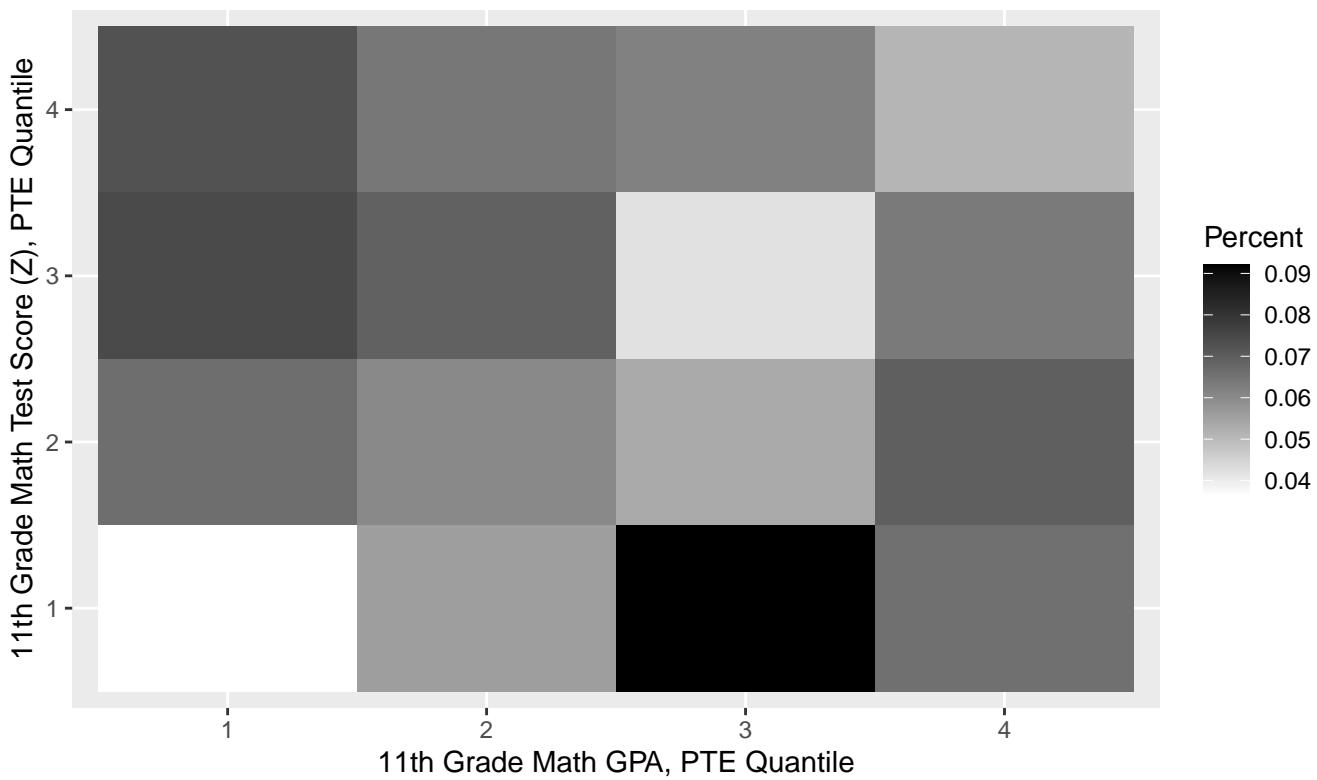
PTE Scatterplot: 11th Grade Math Test Score (Z) vs Graduated on-time



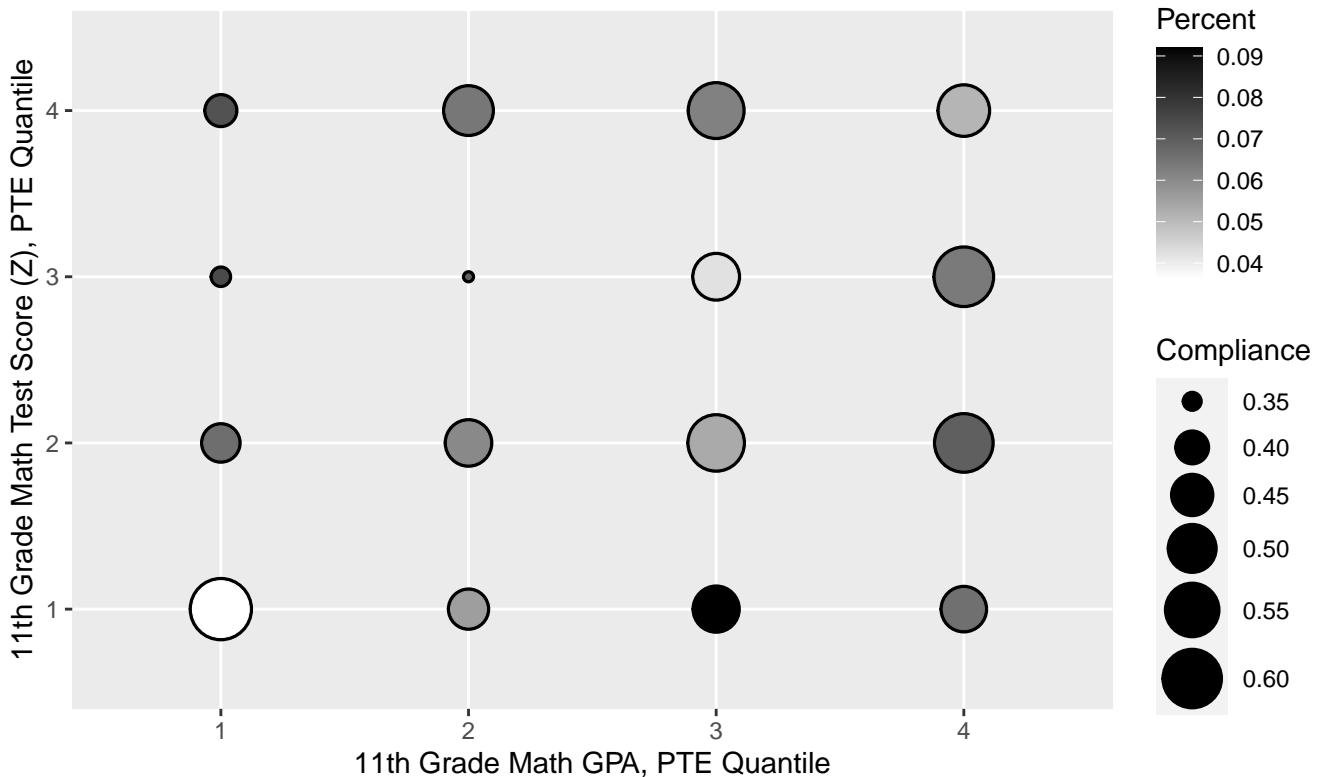
PTE Densities: 11th Grade Math Test Score (Z) vs Graduated on-time



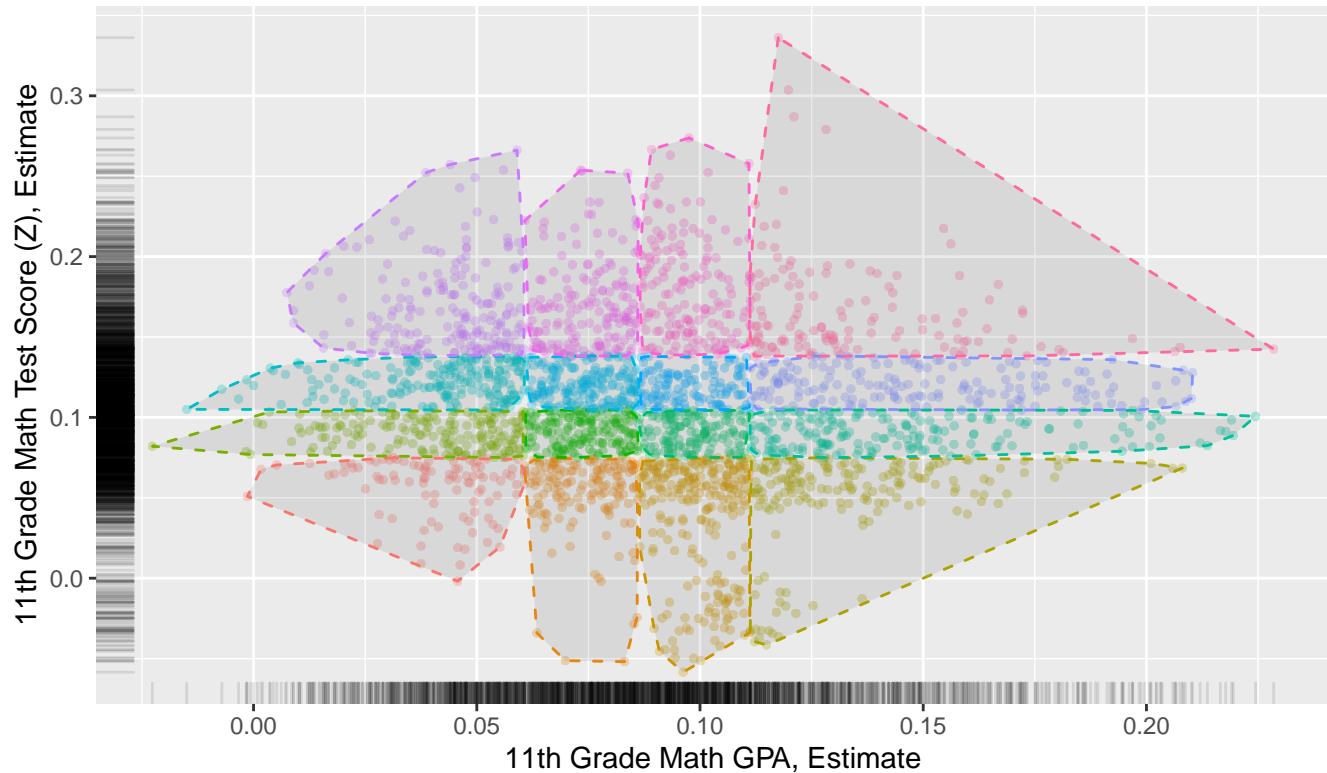
PTE Decile Plot: 11th Grade Math Test Score (Z) vs. 11th Grade Math GPA



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

