Appendix C to the 2015 Colorado Growth Model Report: Investigation of Potential Ceiling and Floor Effects.

Adam R. VanIwaarden
Damian W. Betebenner
National Center for the Improvement of Educational Assessment (NCIEA)

April 2016

1 Introduction

In the 2014-2015 academic year, Colorado transitioned from the Transitional Colorado Assessment Program (TCAP) to the Colorado Measures of Academic Success (CMAS). The transition included numerous changes to the assessment system including the incorporation of new performance standards. As other states have gone through similar assessment transitions, many have observed ceiling and floor effects in the new assessments (i.e. a relatively large proportion of students scoring at/near the scale extremes). This has occurred despite purported improvements in assessment qualities that should prevent these effects (e.g. adaptive tests). Regardless of the source of assessment ceilings/floors, they can telegraph onto the Student Growth Percentile (SGP) calculations making these estimates questionable.

Although very similar in nature, ceiling effects are somewhat more problematic than floor effects because students that consistently achieve the highest scores receive lower than expected growth percentiles and are therefore negatively impacted. Conversely, the consistently lowest achieving students have higher estimated SGPs than would be expected. Although this could be interpreted as a positive impact for these students by giving them higher SGPs, it can also conceal unacceptably low growth that needs to be identified and addressed.

Essentially these problems are caused by the way in which a "percentile" is defined to begin with, and the inability of the assessments (and therefore the SGP model) to make granular distinctions between kids who score at the extremes of the test year after year. As an example, if a group of students were tested with a relatively easy test and 20% of the students had a perfect score, these students would be defined as being in the 80^{th} percentile because they scored higher than 80% of their peers. This is somewhat misleading however, because their score was *equal to* or greater than 100% of their peers and so could be also described as achieving at the 99.99^{th} percentile under an equally valid definition.

To extend this heuristic from achievement to growth, if 50% of those top students also scored perfectly on the next test, we might estimate that they had 50^{th} percentile growth. Although there is nothing *technically* incorrect about this estimate because their growth is fairly typical for their academic peer group, it is an inadequate or unsatisfactory assessment of their growth because they have consistently attained at the highest observable level. Furthermore, if it is typical for their peers to maintain perfect scores then even small deviations from a perfect score could produce *low growth* SGP estimates. Ideally, the expectation is that the majority of SGPs for students scoring at or near the lowest obtainable scale score (LOSS) will be low (preferably less than 5 and not higher than 10), and that SGPs for students with the highest obtainable scale score (HOSS) will be high (higher than 95 and not less than 90).

Given these impacts and the difficulty in detecting them using traditional SGP diagnostic tools, the Center for Assessment has added "Ceiling/Test Effects" indicators to the SGP model goodness of fit plots and is providing all clients even more rigorous diagnostic and descriptive analyses through this Appendix to the annual technical report. This report includes:

1. Plots of the scale score distributions for the current and prior years' tests, which may provide an indication of whether a ceiling or floor is present in either (or both) the current or historical data.
2. Box plots showing the range and distribution of SGPs for *only* the highest and lowest achieving students in the current year.

2 Prior and Current Year Score Distributions

The marginal and conditional distributions of scale scores can serve as a preliminary indicator of potential ceiling or floor effects in the calculation of student growth percentiles. Some minor problems could present themselves if these characteristics are present in either prior or current year scores, and are particularly likely when present in both. The plots below depict distributions for the current year and the most recent prior year used in the SGP calculations. The marginal (individual or univariate) distributions for each year are shown in the first subsections below, followed by the conditional (joint) distributions. Each content area plots are disaggregated by grade.

2.1 Marginal Distributions

Generally there is little evidence of floor/ceiling effects in the TCAP prior scores. When present, these effects appear either as sharp spikes at the extremes of the distributions or as fat, truncated tails.

The CMAS scores are truncated at the floor in some grades and end-of-course test (EOCT) subjects. This could potentially translate to floor effects in growth analyses. No evidence of ceiling effects are present in the marginal distributions alone.

2.1.1 ELA

Fig. C.1: Marginal distributions of prior scale scores: Reading.

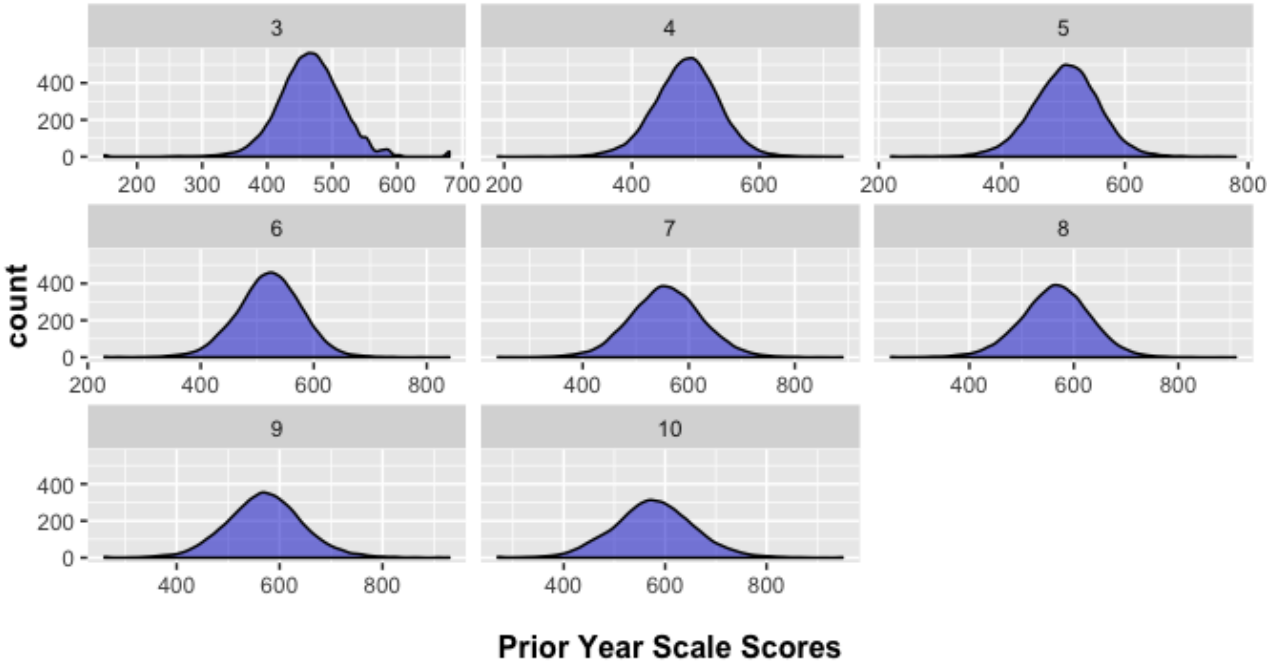


Fig. C.2: Marginal distributions of prior scale scores: Writing

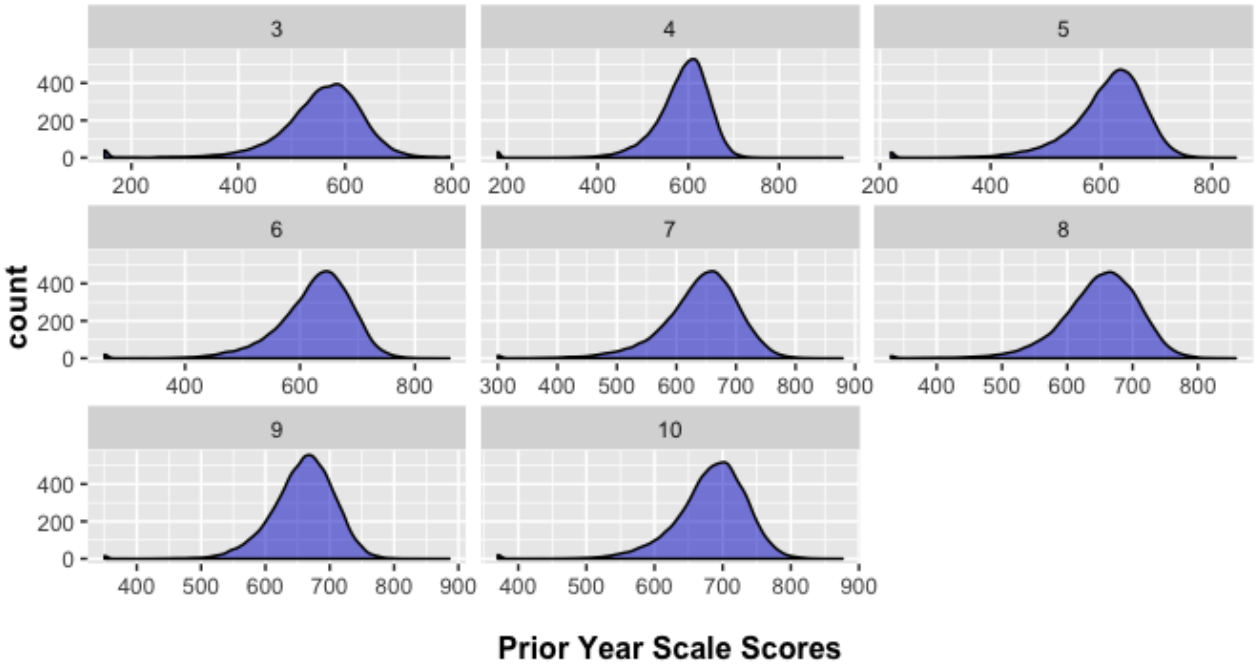
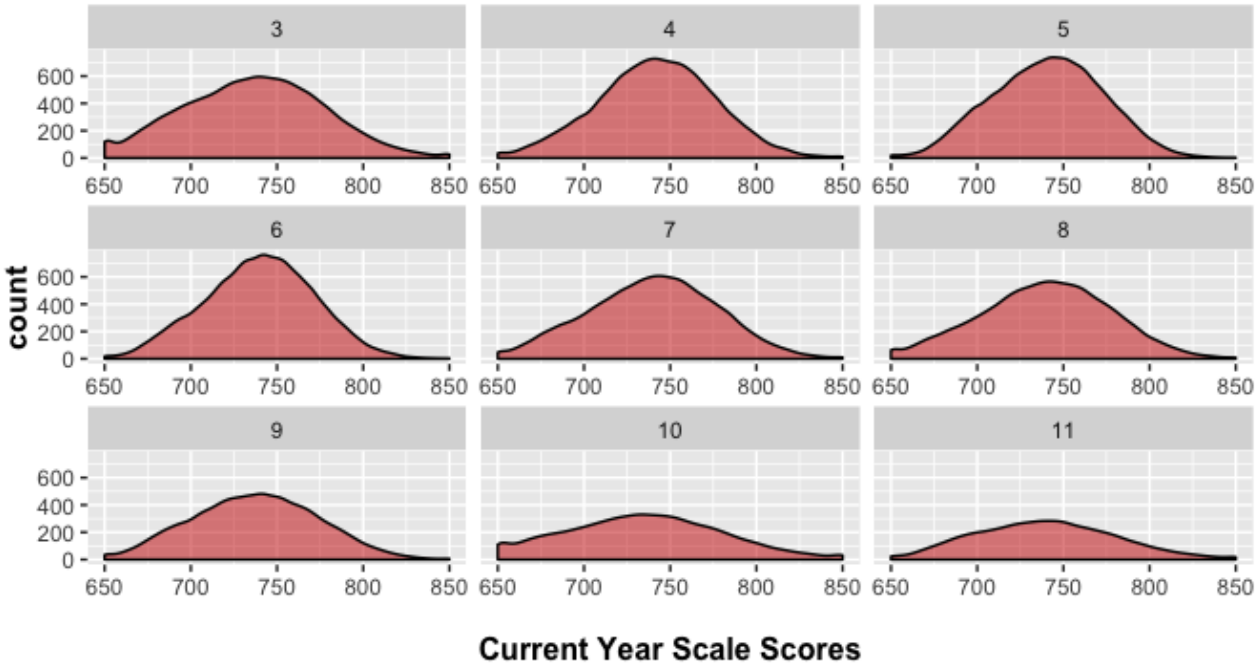


Fig. C.3: Marginal distributions of current scale scores: ELA.



2.1.2 Mathematics

Fig. C.4: Marginal distributions of prior scale scores: Mathematics.

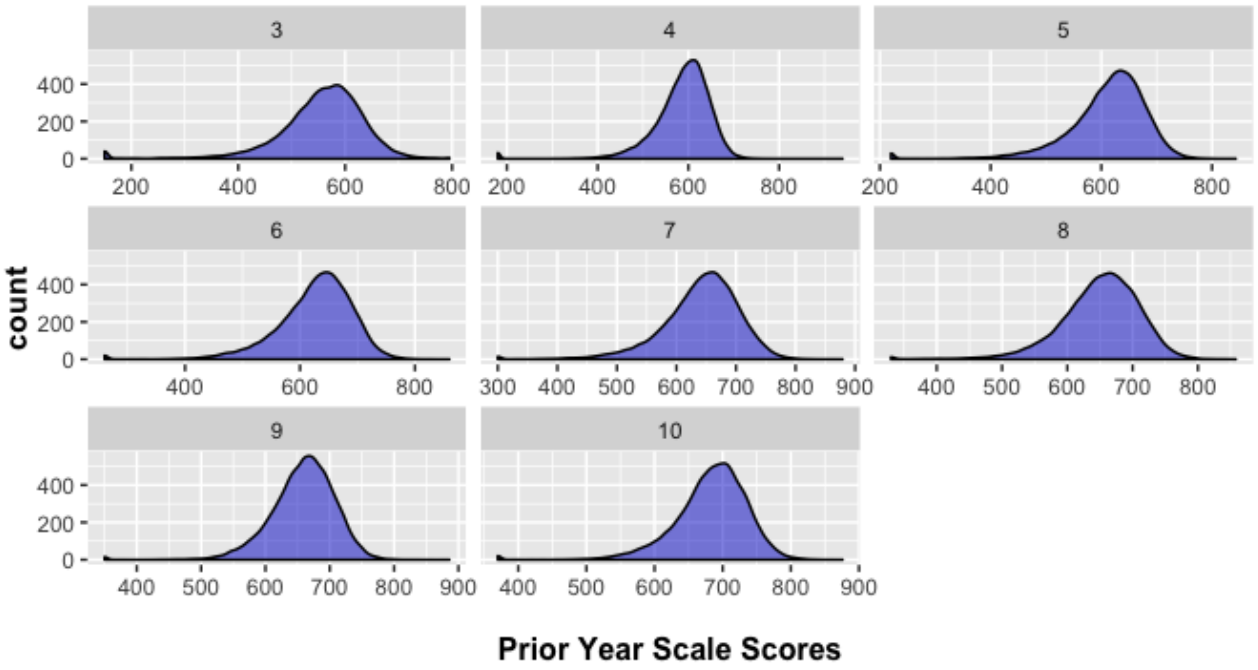
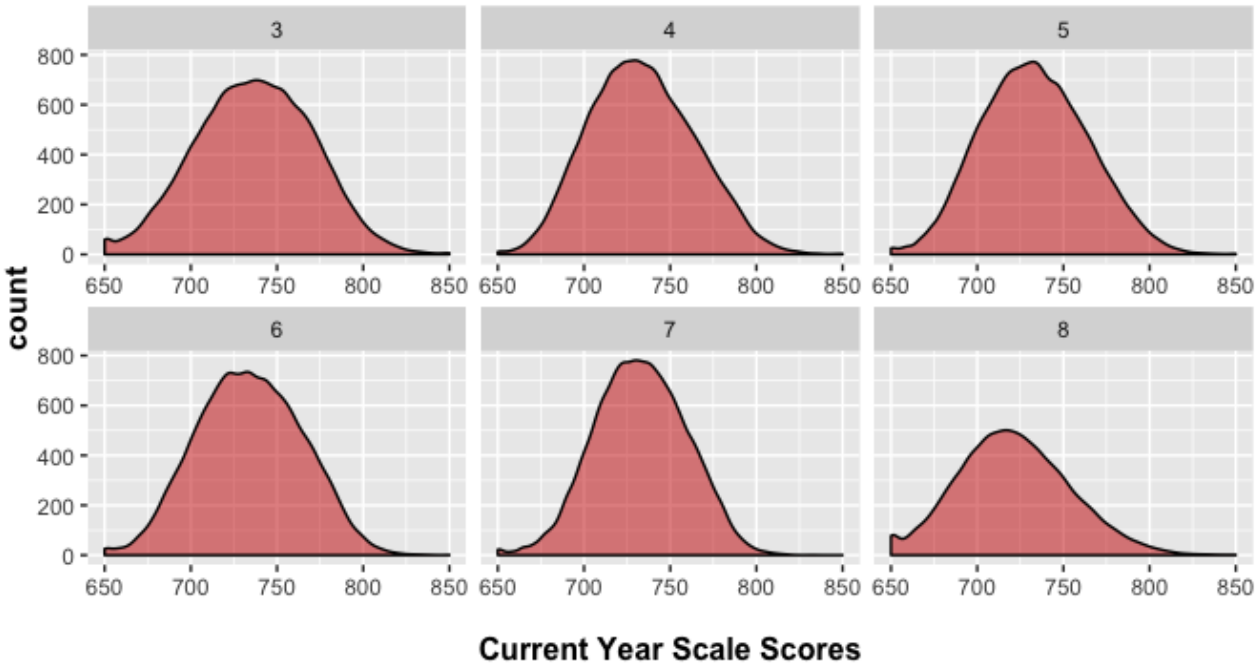


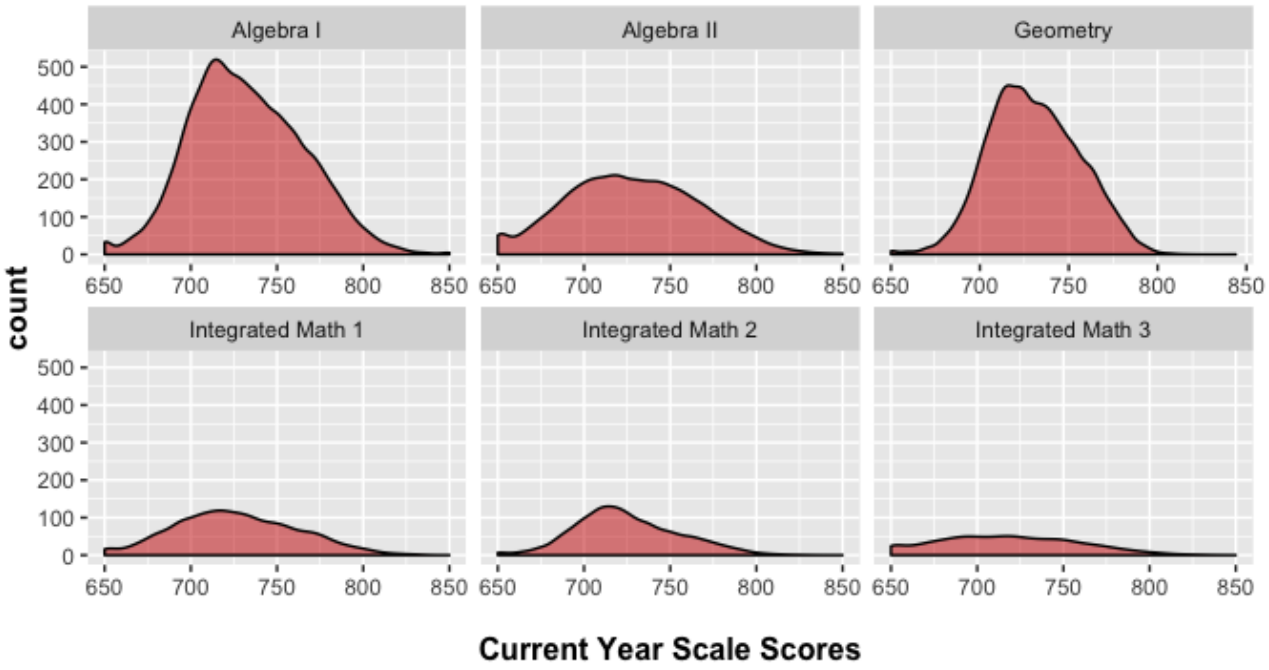
Fig. C.5: Marginal distributions of current scale scores: Mathematics.



2.1.3 End-of-Course Test Subjects

Only the current year distributions are presented here because this is the first year of CMAS testing in EOCT subjects in Colorado.

Fig. C.6: Marginal distributions of current scale scores: End-of-Course Subjects.



2.2 Conditional Distributions

The marginal density plots provide a limited amount of information, particularly for the potential for ceiling/floor effects in the calculation of ***growth***. In order to provide a more nuanced view of the relationship between the prior and current scale scores, the following plots depict the conditional (joint) distributions for each content area and grade level. These plots start with a basic scatter plot of each student's scores, and on top of this is layered 1) **green contour lines** to provide a sense of the joint density, 2) a **magenta non-linear line** identifying the bivariate relationship between prior and current scores, and 3) **rug plots** that describe the marginal distributions (as above, the prior scores are blue and current scores are red).

Ceiling or floor effects often appear as dark shaded points in the extreme top-right or bottom-left corners of the plots. This suggests that staying at the extremes from year to year is not uncommon, which may lead to odd growth estimates for these chronically high/low achieving students. For the 2014 and 2015 Colorado data, we begin to see concerning indicators in most content area and grade combinations. Along with the tell-tale points in the extremes, we also see situations in some EOCT course progressions where data is sparse in the upper extremes. In these cases there can also be ceiling impacts for the one or two outliers with consistently high achievement (see for example the bottom left panel of Figure C.11 for Algebra I with 9^{th} grade math priors).

2.2.1 ELA

Fig. C.7: Conditional distribution(s) of current and prior scale scores: ELA.

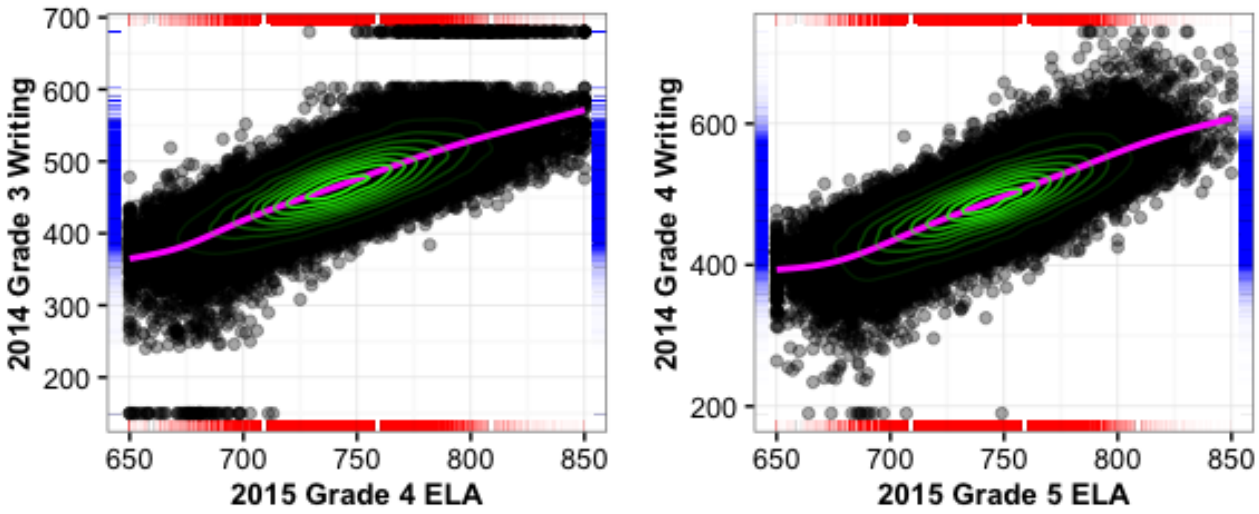
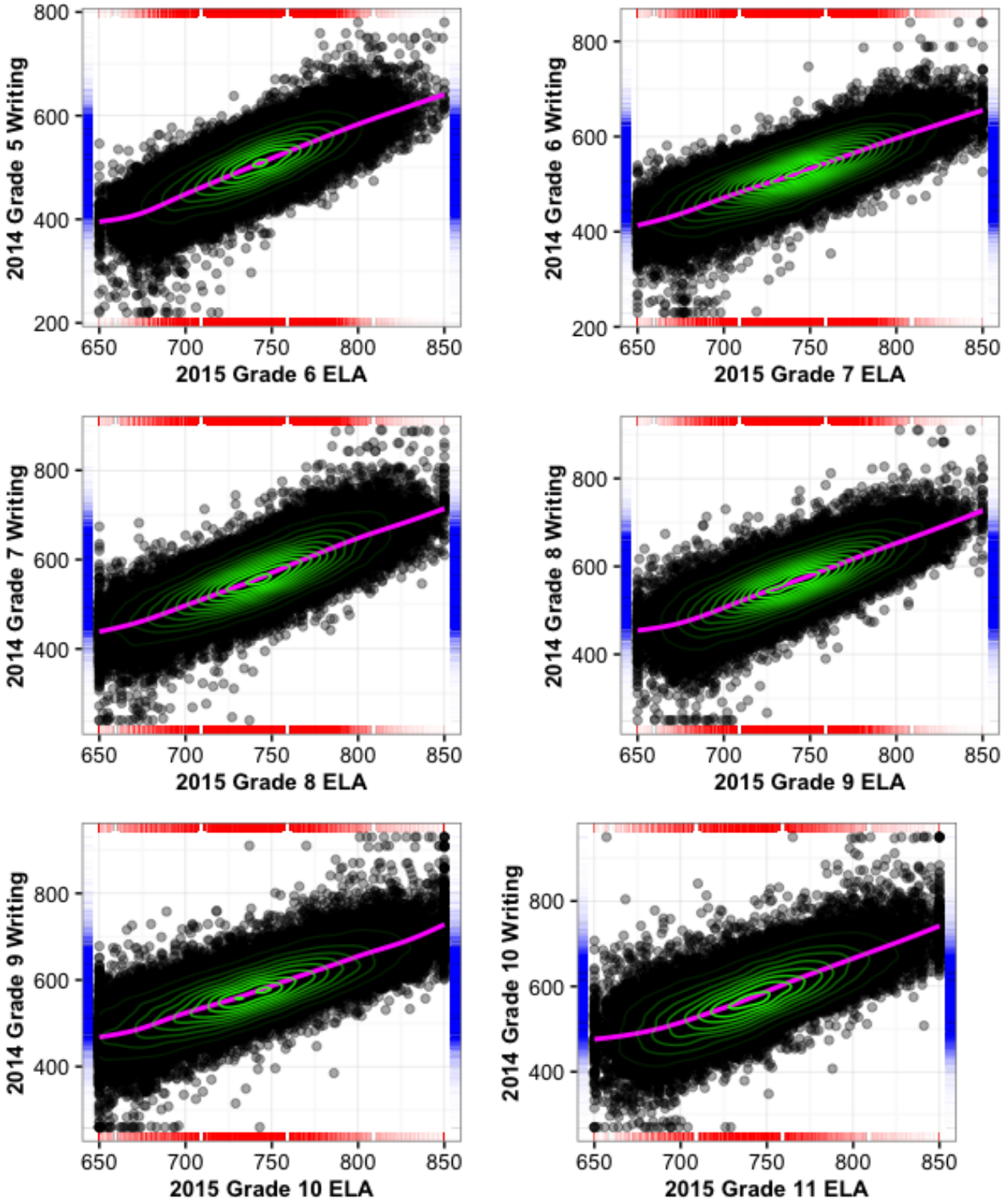
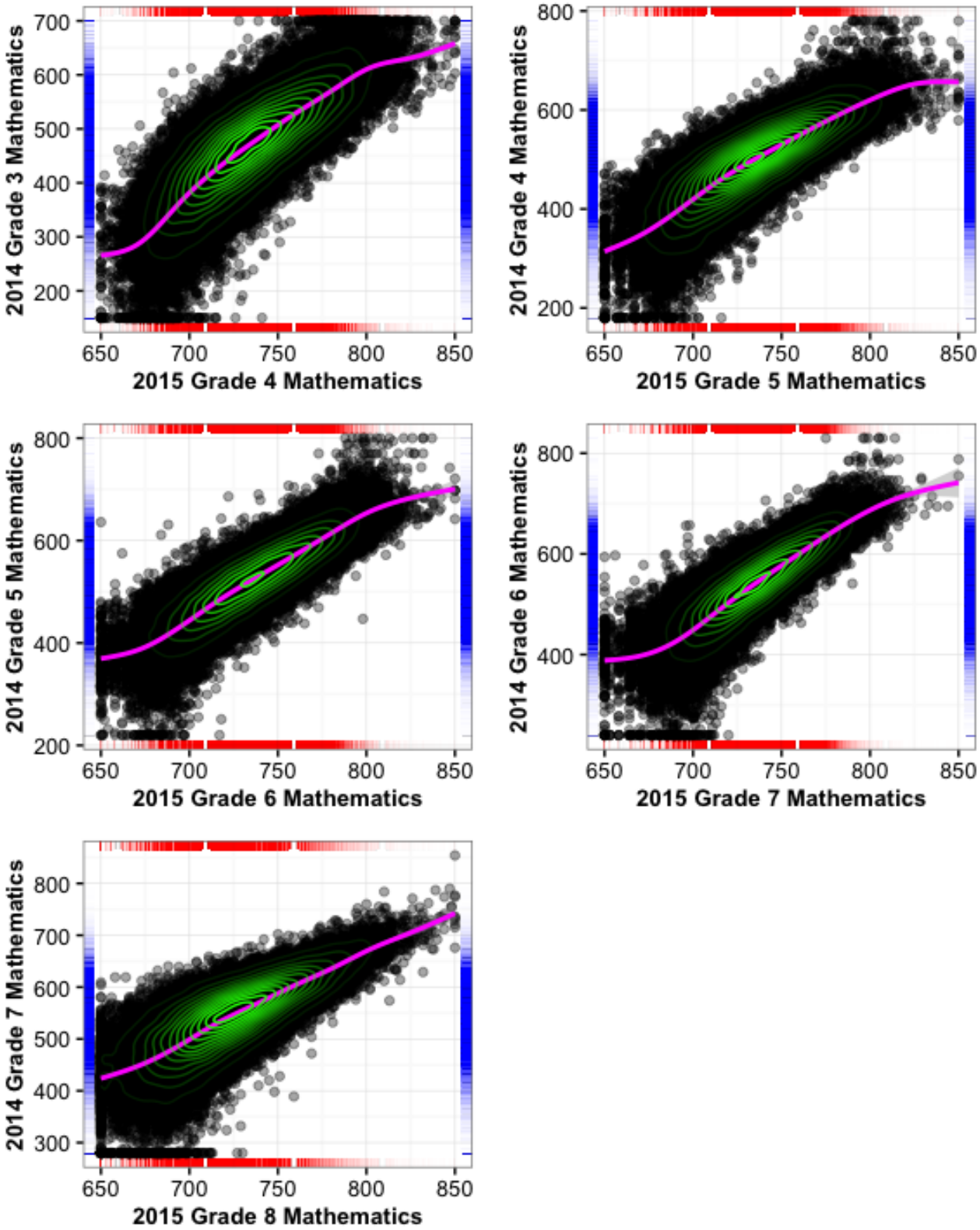


Fig. C.8: Conditional distribution(s) of current and prior scale scores: ELA *Continued.*



2.2.2 Mathematics

Fig. C.9: Conditional distribution(s) of current and prior scale scores: Mathematics.



2.2.3 End-of-Course Test Subjects

The conditional density plots for the EOCT subjects are displayed below. The most recent prior is used in each plot to provide insight on each academic peer (norm) group that were analyzed separately. Note the bottom left plot in Figure C.11 below for Algebra I, particularly the point in the upper-right hand area of the plot. This is a potential indicator for a ceiling effect issue. Several other individual plots in other subjects also display a similar indicators.

Fig. C.11: Conditional distribution(s) of current and prior scale scores: Algebra I.

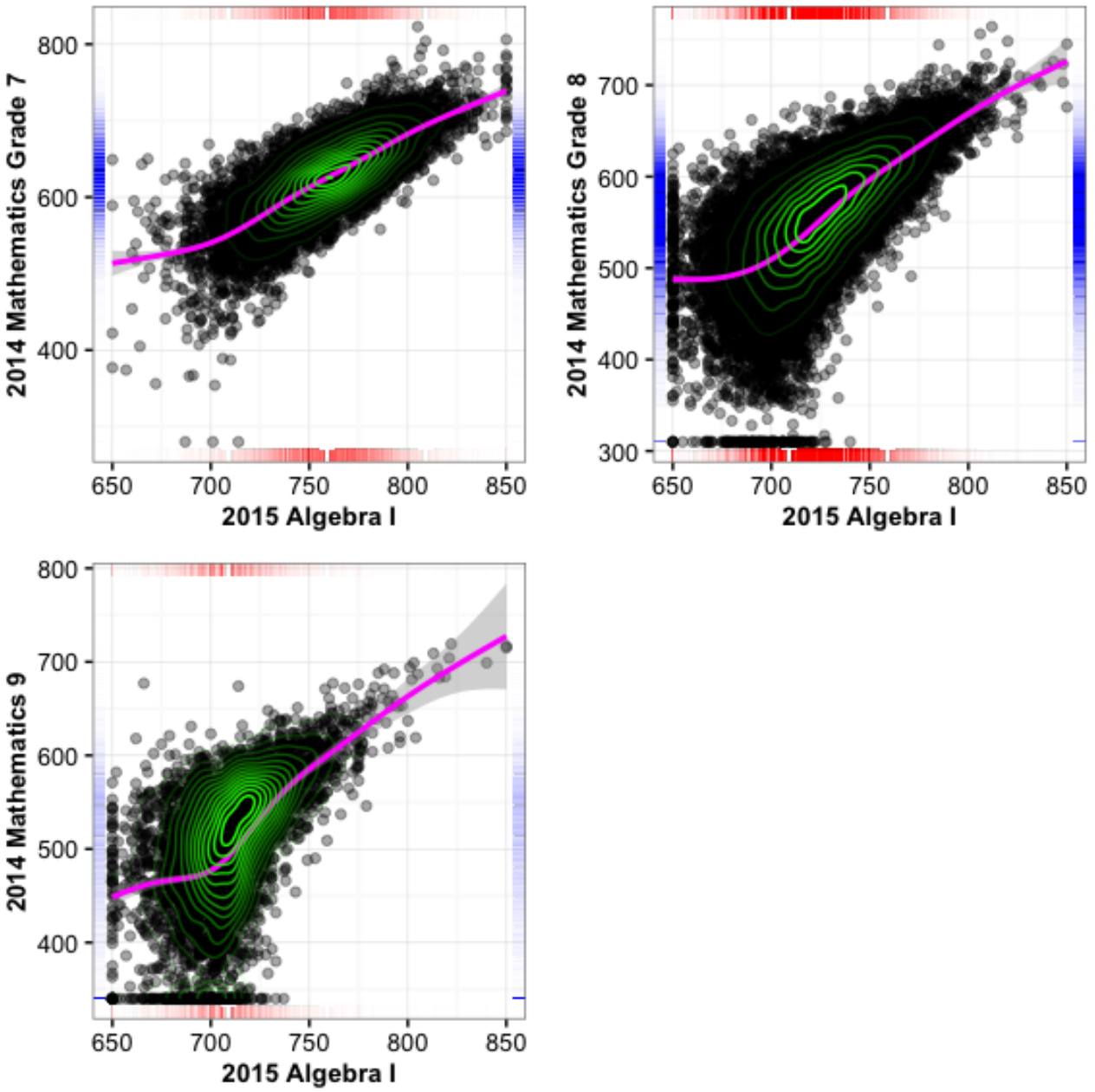


Fig. C.12: Conditional distribution(s) of current and prior scale scores: Algebra II.

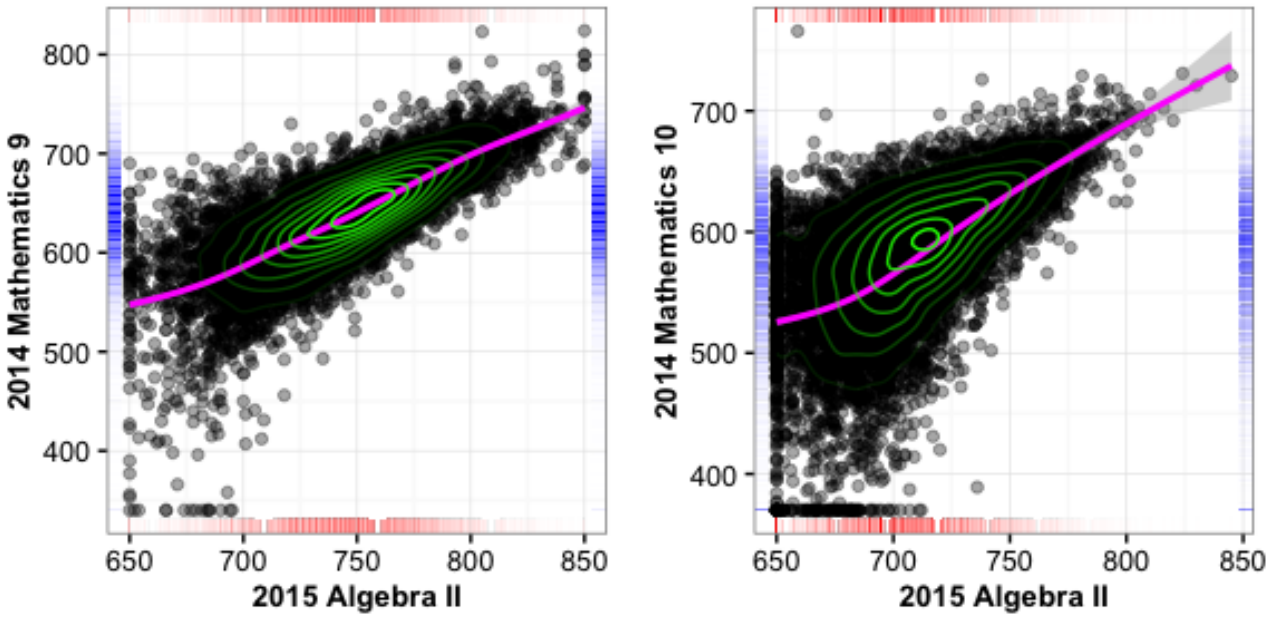


Fig. C.13: Conditional distribution(s) of current and prior scale scores: Geometry.

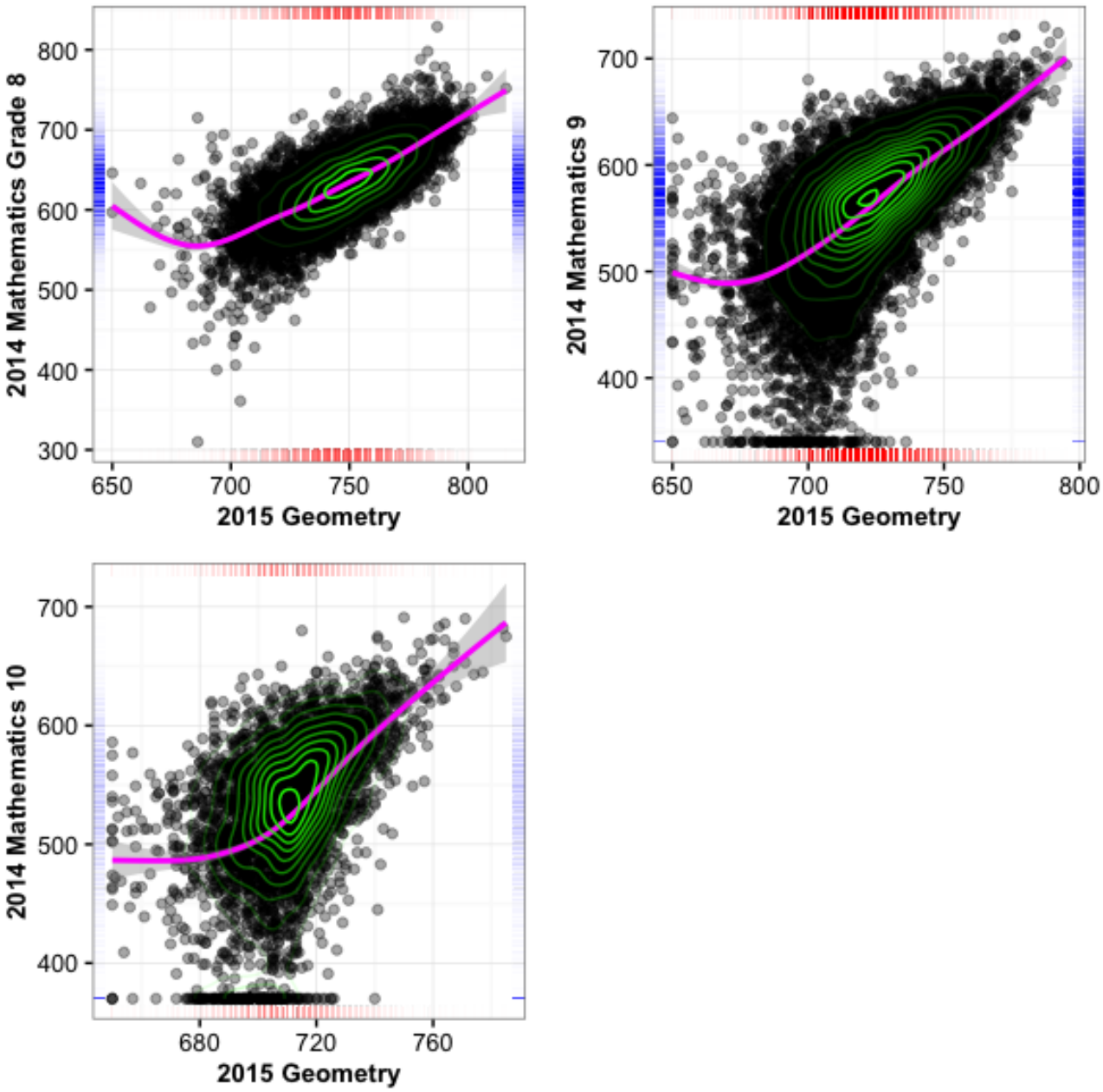


Fig. C.14: Conditional distribution(s) of current and prior scale scores: Integrated Math 1.

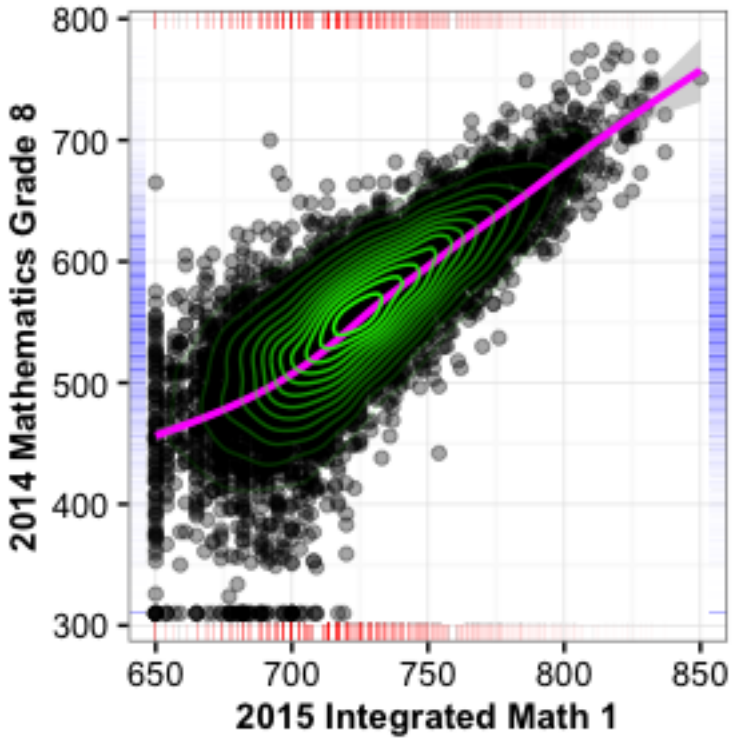


Fig. C.15: Conditional distribution(s) of current and prior scale scores: Integrated Math 2.

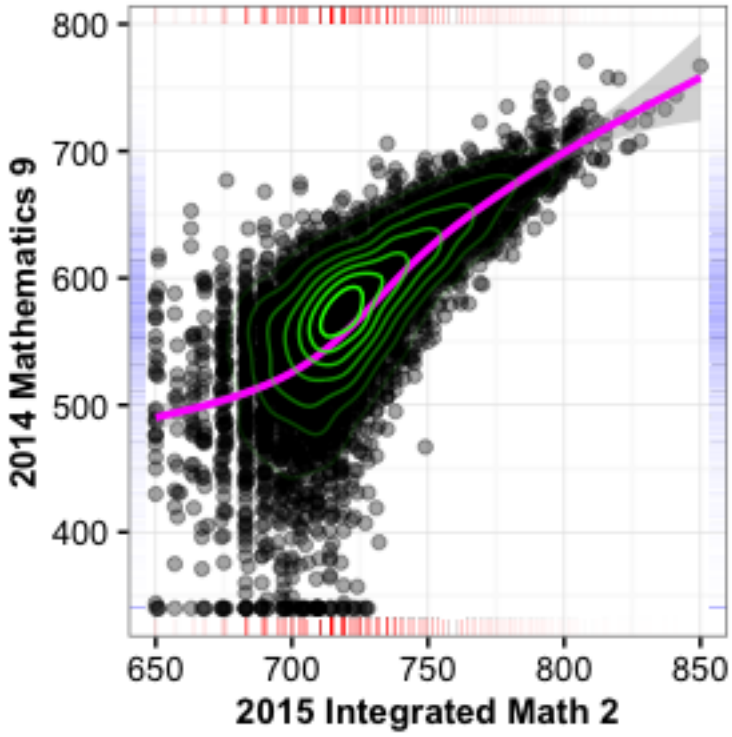
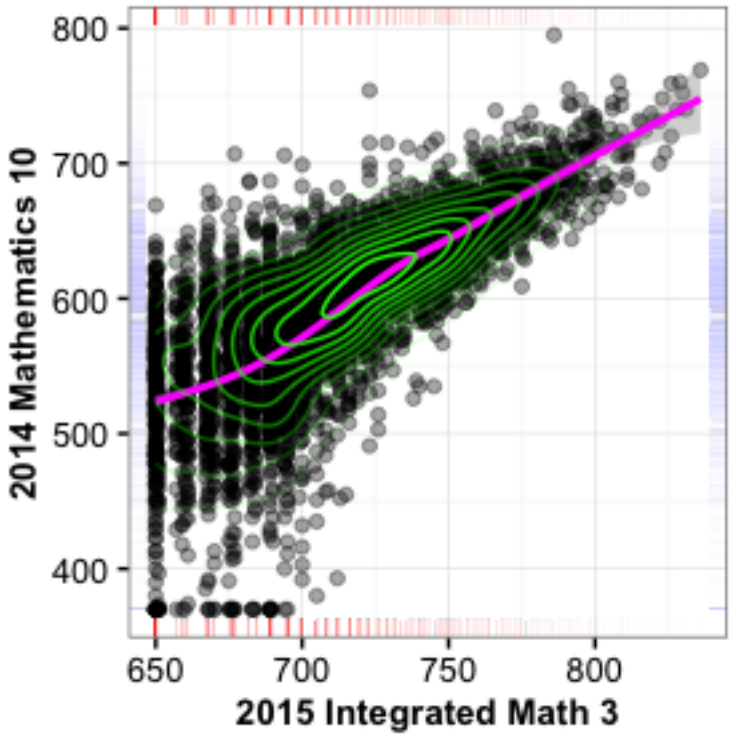


Fig. C.16: Conditional distribution(s) of current and prior scale scores: Integrated Math 3.



3 SGP Ranges for the Highest and Lowest Achieving Students

In order to isolate the impact of assessment ceilings/floors on student growth percentile (SGP) calculations, the following section provides box plots of the distribution of SGPs for the highest and lowest achieving students. We are specifically interested in the SGPs for students scoring at the highest/lowest obtainable scale score (HOSS/LOSS - i.e. the actual ceiling/floor). However, in order to assure that an adequate number of students are included, the first set of plots in each subsection uses, at a *minimum*, the highest/lowest 50 scores. Note that this roughly corresponds to the number of students used in the SGP model goodness of fit plots, and this is why these plots are provided as a starting point for this part of the investigation. All students with a score in these students' range of scores are included. Consequently, the number of students in each box plot may be greater than 50 (the exact number is shown at the margins in red text).

The second set of box plots in the following subsections isolate ***only*** those students scoring the HOSS/LOSS. These plots may then incorporate a widely varying number of students depending on the prevalence of a ceiling/floor in the current year.

The box plots provide several descriptive statistics. The dark line within the box marks the *median* SGP, while the ends ("hinges") of the boxes correspond to the first and third quartiles (the 25^{th} and 75^{th} percentiles). The upper whisker extends from the hinge to the highest value that is within 1.5 $\times$ IQR of the hinge, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the hinge to the lowest value within 1.5 $\times$ IQR of the hinge. Data beyond the end of the whiskers are outliers and plotted as points. Evidence of a *lack* of either a ceiling or floor effect would be to have all high achieving students with SGPs near 99 and all low achieving students with SGPs near 1. That is, the desired visual evidence is a solid line at SGP = 99/1.

The grade level results show considerable ceiling effects for ELA as well as 8^{th} grade mathematics with many students' SGPs well below 90. There is also evidence of floor effects in these content areas, although this is not as strong as the ceiling (mainly outlier students with an SGP above 10).

3.1 Grade Level Content Areas

Fig. C.17: SGP distributions for highest and lowest 50+ scale scores by content area and grade level.

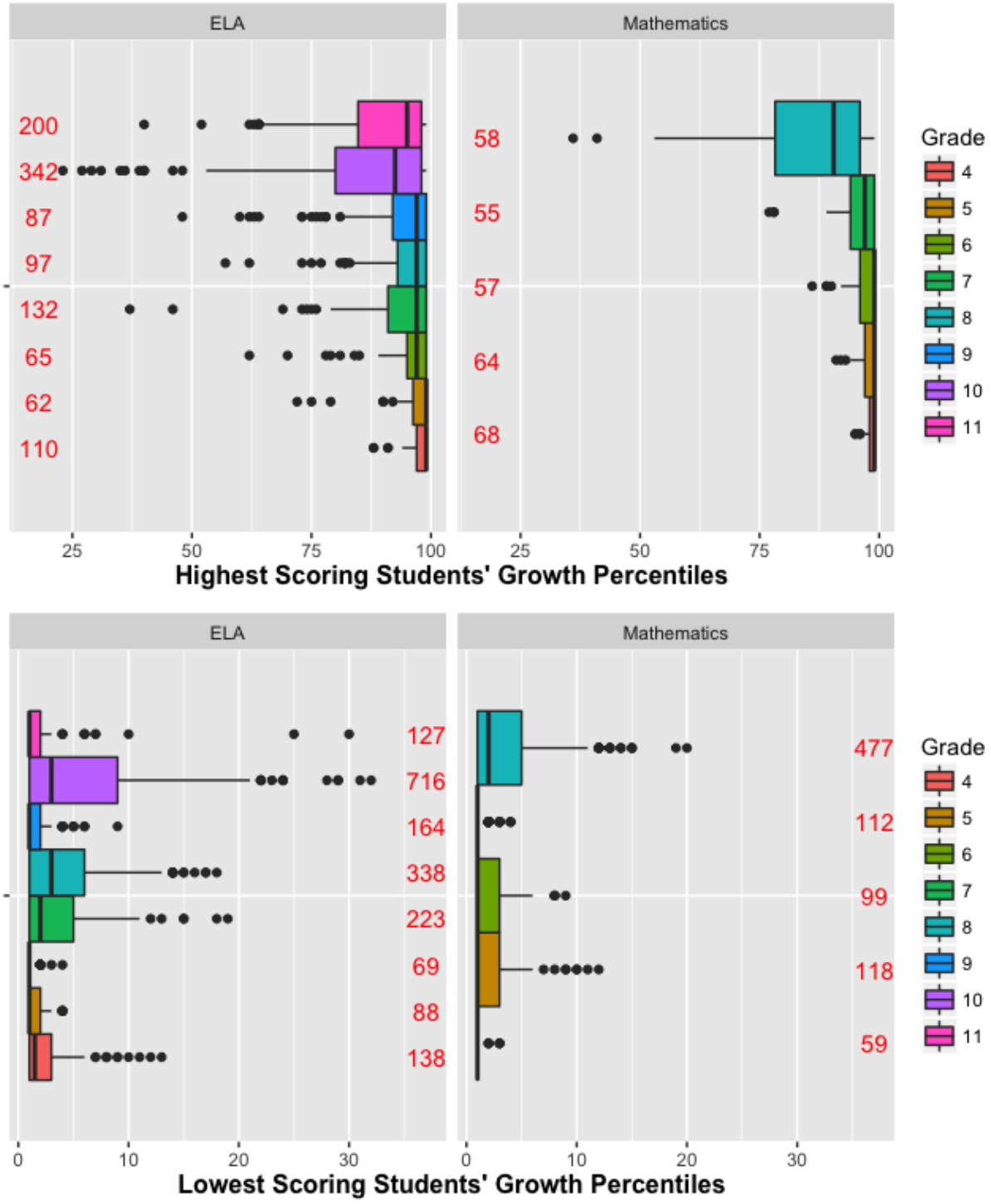
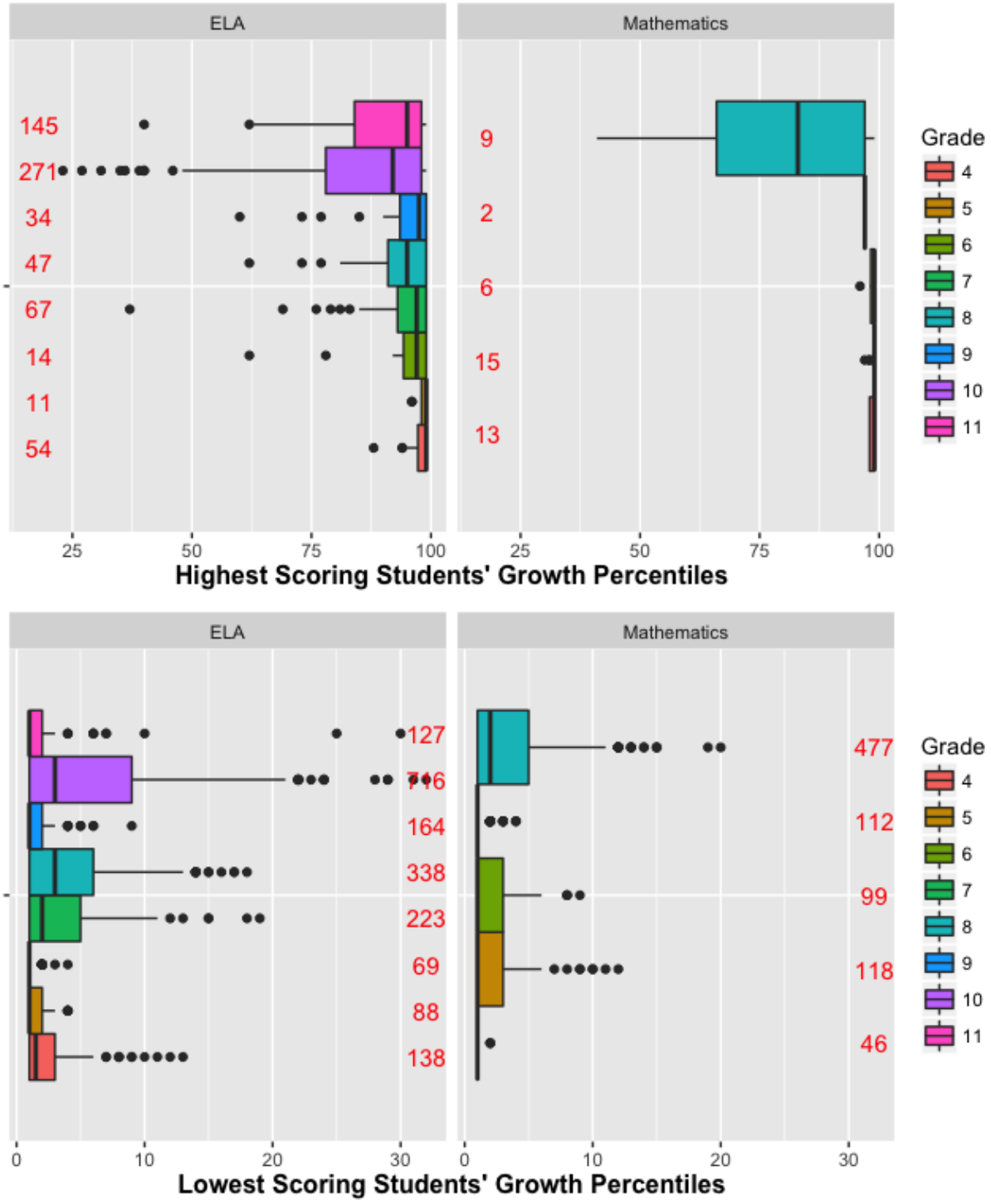


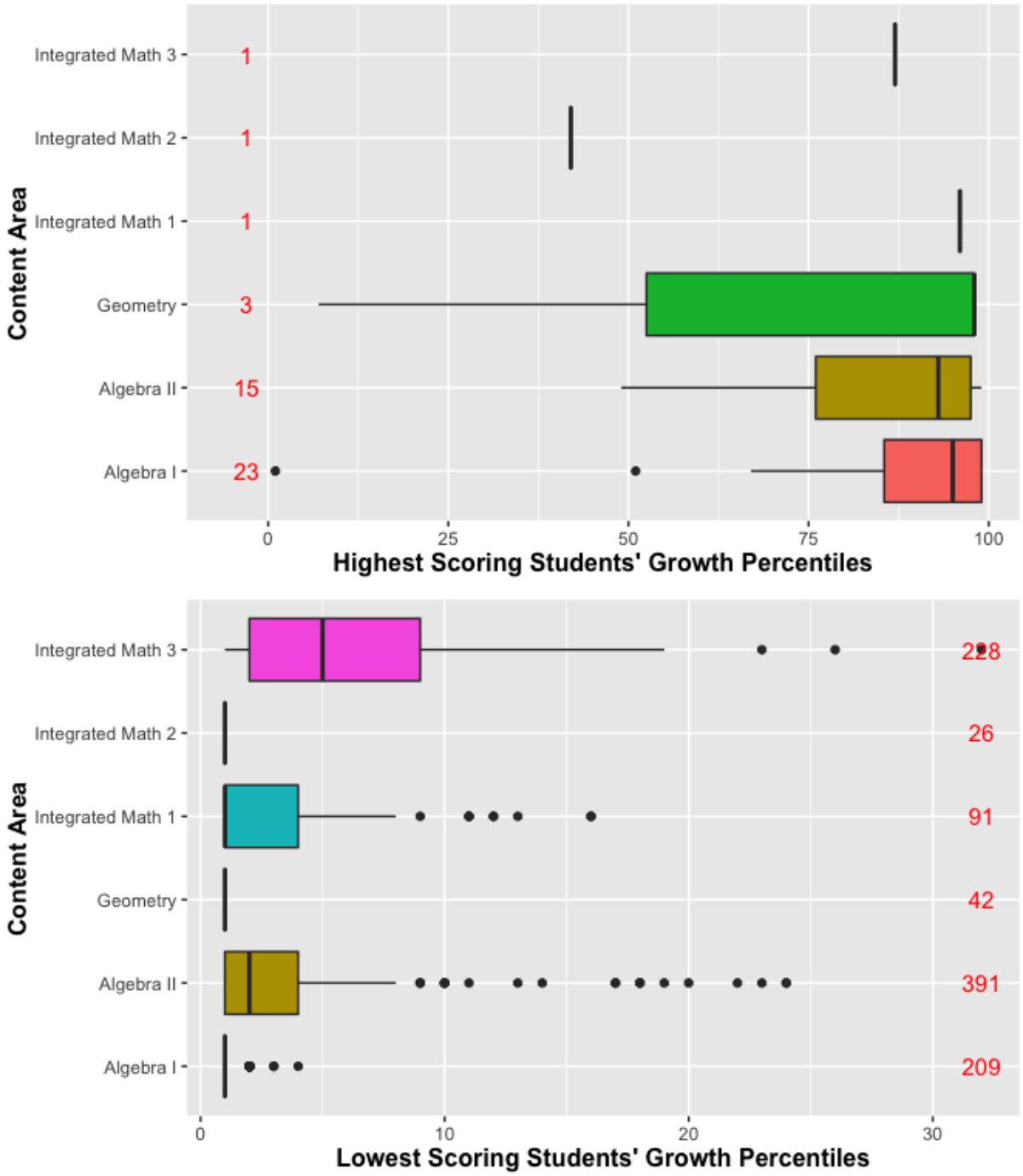
Fig. C.18: SGP distributions for the HOSS and LOSS scores by content area and grade level.



3.2 EOCT Content Areas

The end-of-course subject results are shown here for *only* the LOSS and HOSS students. There are several subjects for which potential ceiling and floor effects are evident. These subjects are disaggregated further in order to more appropriately address these concerns.

Fig. C.19: EOCT SGP distributions for the HOSS and LOSS scores by content area.



The end-of-course box plots can be disaggregated further by the most recent prior (i.e. their primary constituent norm groups). The following box plots disaggregate each EOCT subject by norm groups. The most serious ceiling impact can be seen in the Algebra I progression with 9^{th} grade mathematics as the most recent prior. Of the two students scoring the HOSS, one received an SGP of 1 and the other 51. These students were very high achieving outliers in their cohorts (as shown in Figure C.11 above), and consequently the SGP model was unable to make acceptable estimates for their growth.

Fig. C.20: SGP distributions for the HOSS and LOSS scores by norm group: Algebra I.

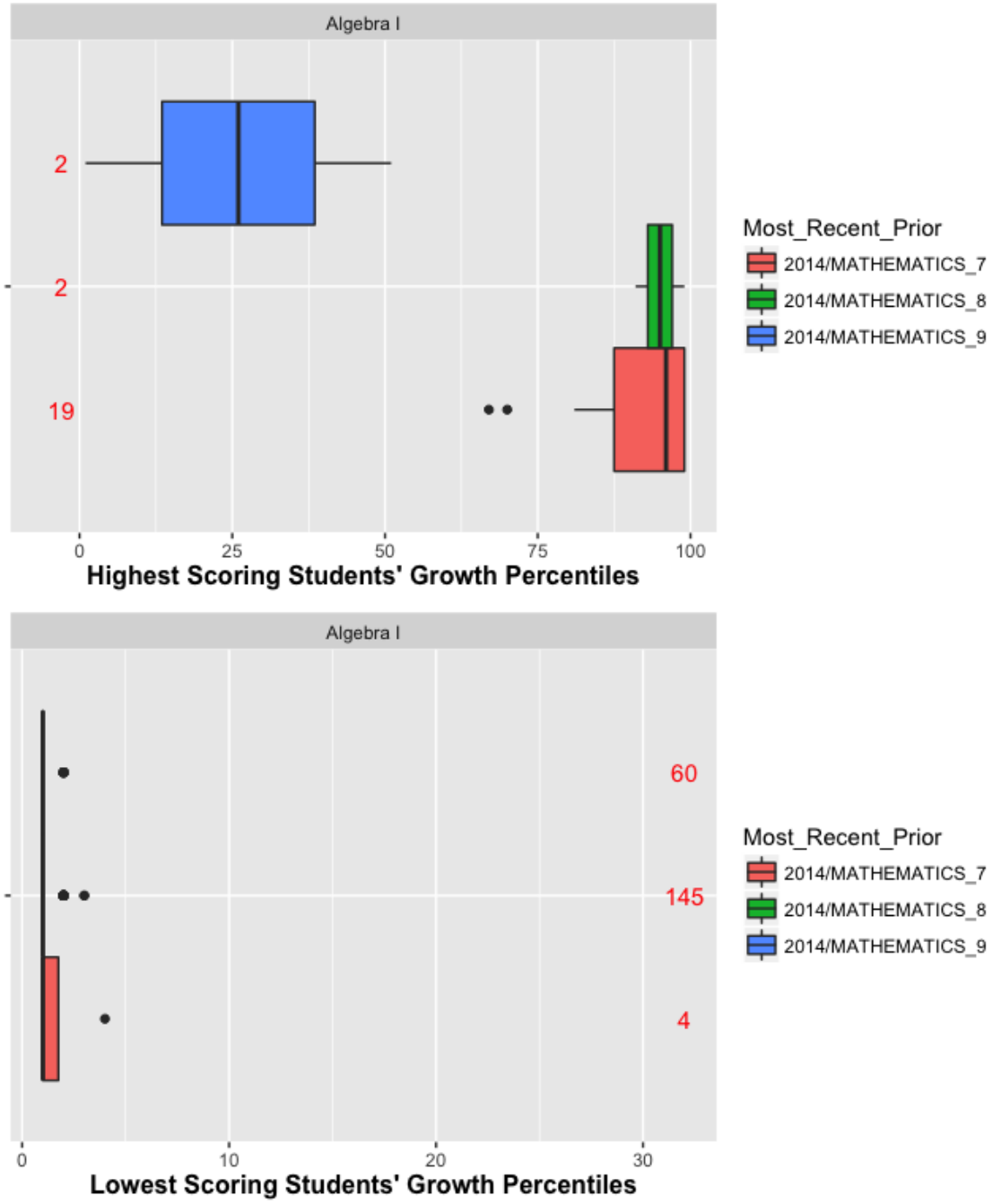


Fig. C.21: SGP distributions for the HOSS and LOSS scores by norm group: Algebra II.

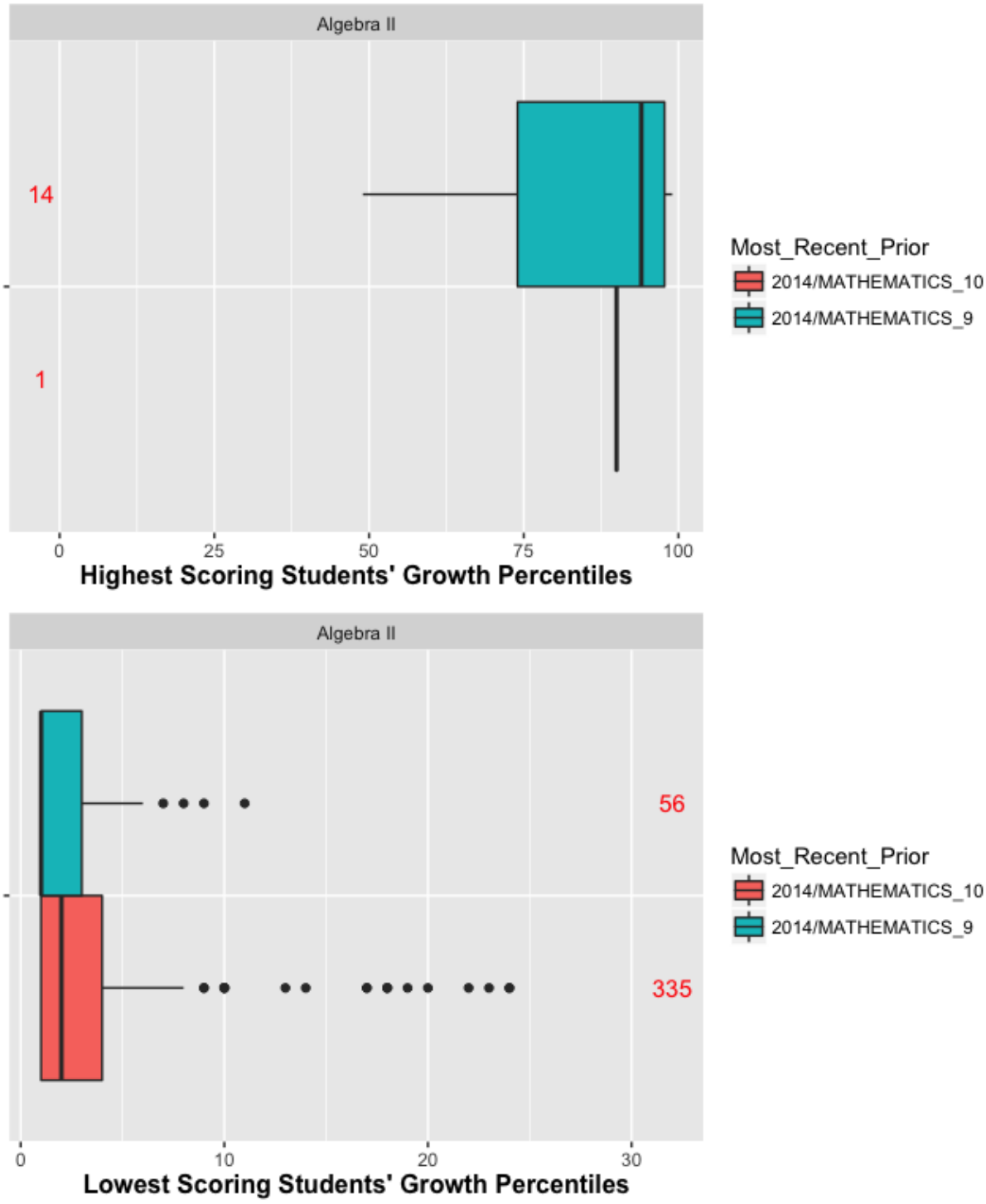


Fig. C.22: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Geometry.

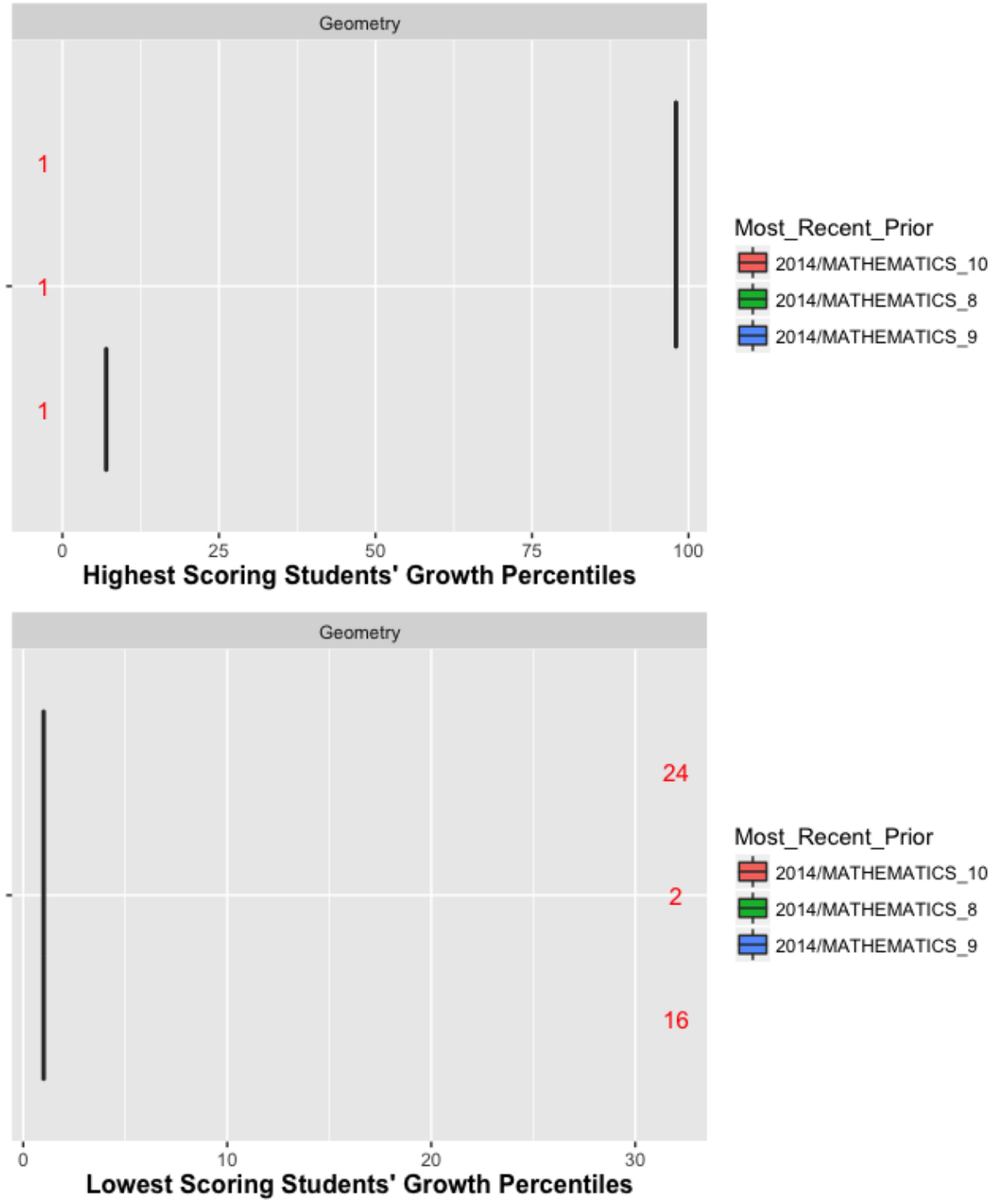


Fig. C.23: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Integrated Math 1.

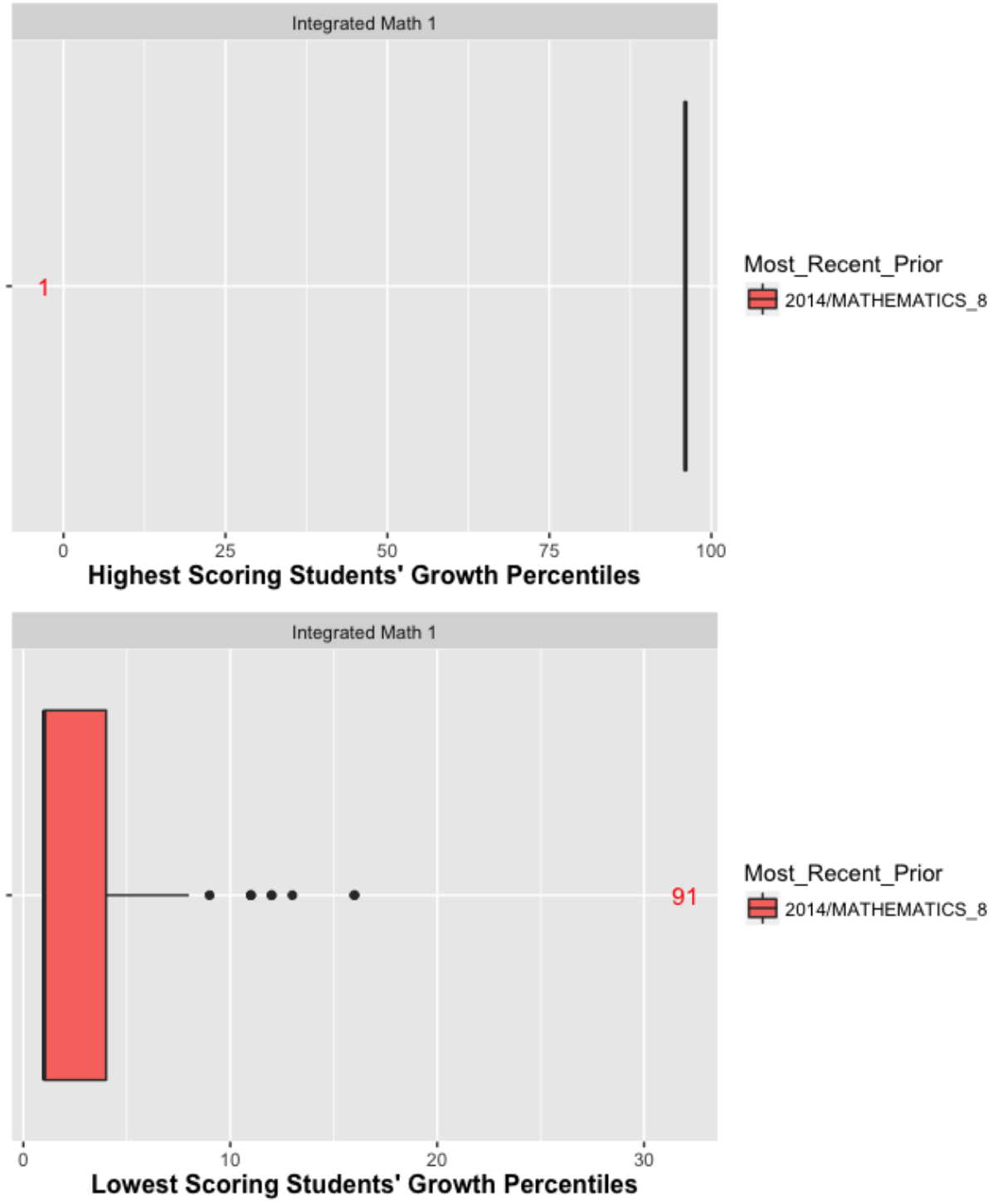


Fig. C.24: EOCT SGP distributions for the HOSS and LOSS scores by norm group: Integrated Math 2.

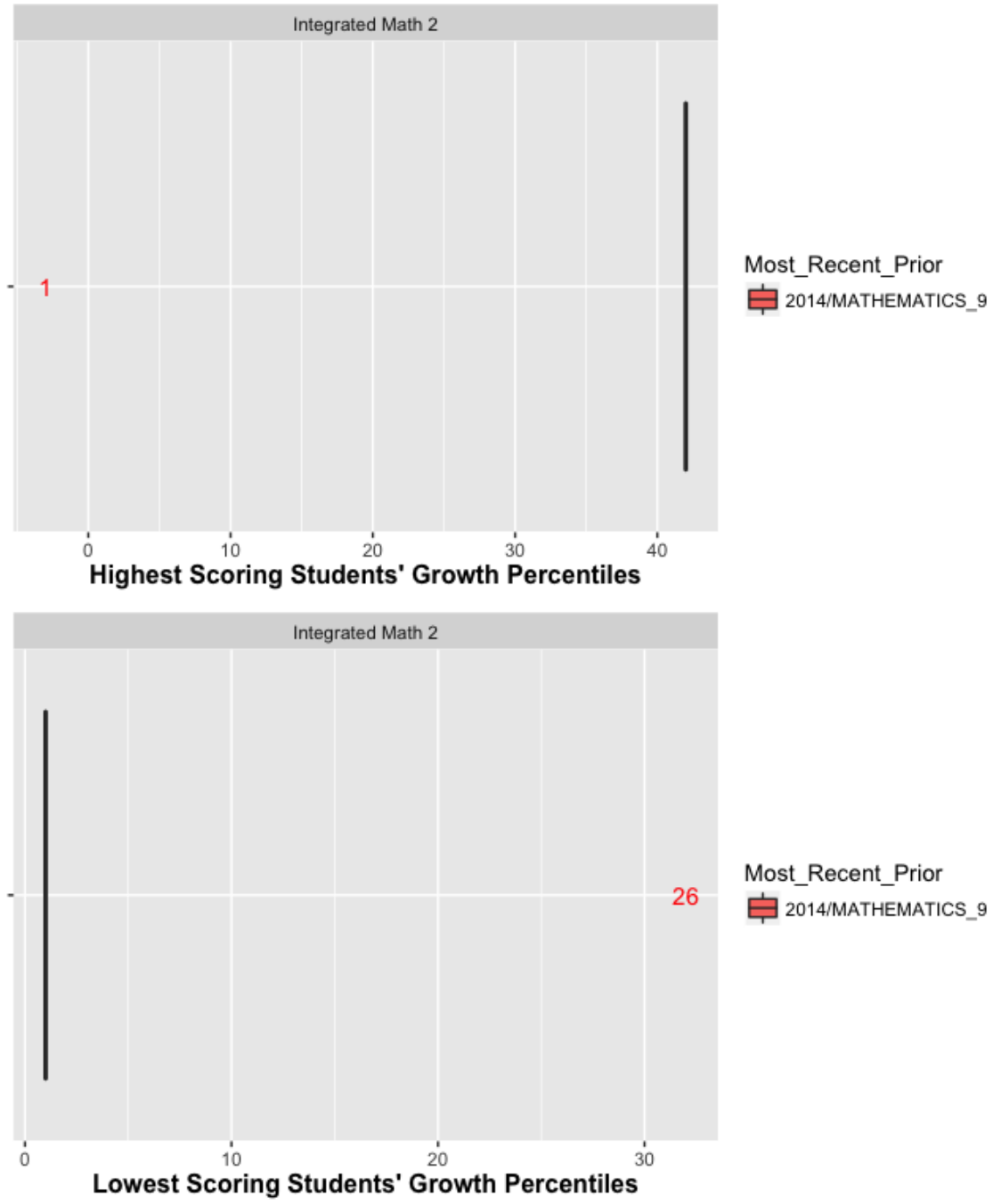
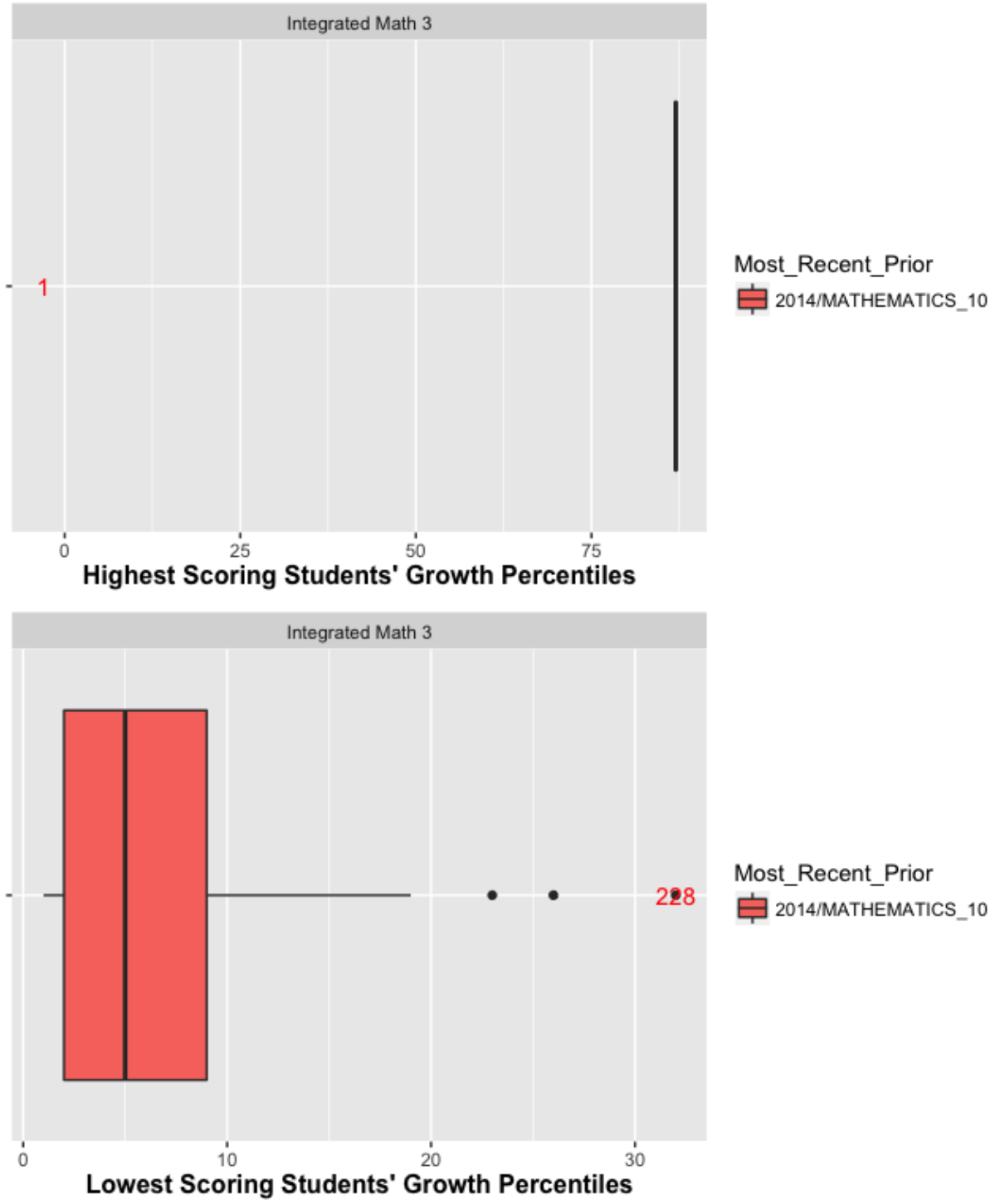


Fig. C.25: EOC SGP distributions for the HOSS and LOSS scores by norm group: Integrated Math 3.



4 Discussion

Overall there is substantial evidence of both floor or ceiling effects in the 2015 Colorado SGP analyses. Both grade level and end-of-course test subjects are impacted to some degree.

When ceiling or floor effects are encountered, there are several ways in which they can be "corrected" manually or analytically. These include (but not limited to):

1. Convert all students scoring at the HOSS (LOSS) to 99 (1).
2. Run SGP analyses with more granular scores. For example, many tests that use Item Response Theory (IRT) to analyse test results provide scaled scores that enforce an artificial ceiling (floor), but also have more granular achievement scores available (IRT θ estimates).
3. Leave the results without a correction.