Informed Data Visualization

Joanne R. Wendelberger, Ph.D.

Statistical Sciences Group, Los Alamos National Laboratory

joanne@lanl.gov

Invitational Summit on Educational Data Visualization, Austin, Texas, May 5-6, 2014

UNCLASSIFIED

Abstract

Data visualization can be a powerful tool for extracting insights from complex data. Modeling and analysis using rigorous statistical methodology can enable informative visualizations for understanding behavior of populations as well as individuals within the populations. An important component of the analysis process is the quantification of uncertainty associated with models, data, estimates, and predictions. With the explosion of data being collected, it will become increasingly important to develop computational infrastructure and tools required for analysis and visualization of large, heterogeneous data in support of research efforts, policy, and informed decision-making.

Introduction

- Data Visualization can be a powerful tool for extracting insights from education data.
- Statistical analysis and modeling enable informative visualization of educational data, addressing individual and population behavior.
- The increasing amount of education data being collected requires advances in information science and technology to support informed data visualization for policy and decision-making.

Acknowledgements

- Dave Higdon, LANL
- Mike Fugate, LANL
- Steve Stringer, LANL
- Ben Sims, LANL
- Stephan Eidenbenz, LANL
- Pat Kelly, LANL
- Michael Strong, UCSC
- Paul Resta, UT-Austin
- Ken Tothero, UT-Austin
- Tom Benton, UT-Austin
- Gordon Freedman, NLET
- Marcy Lauck, NLET
- Jim Ahrens, LANL
- Jon Woodring
- Kary Myers, LANL
- David Rogers, LANL
- Francesca Samsel, UT-Austin
- Greg Abram, UT-Austin

Outline

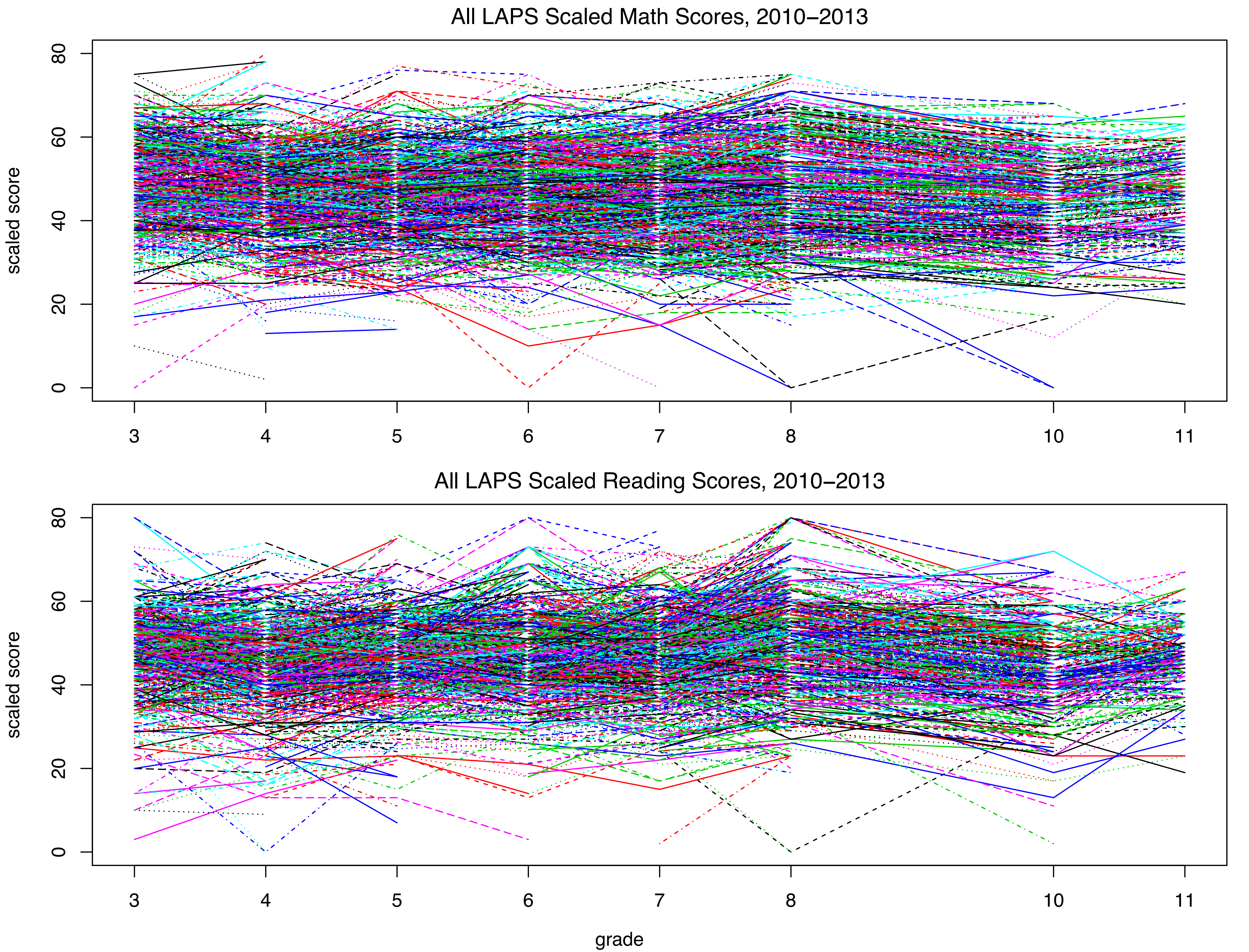
- Statistical Visualization of New Mexico Education Data
- An Agent-Based Approach for Education Modeling
- Stat/CS Visualization Research

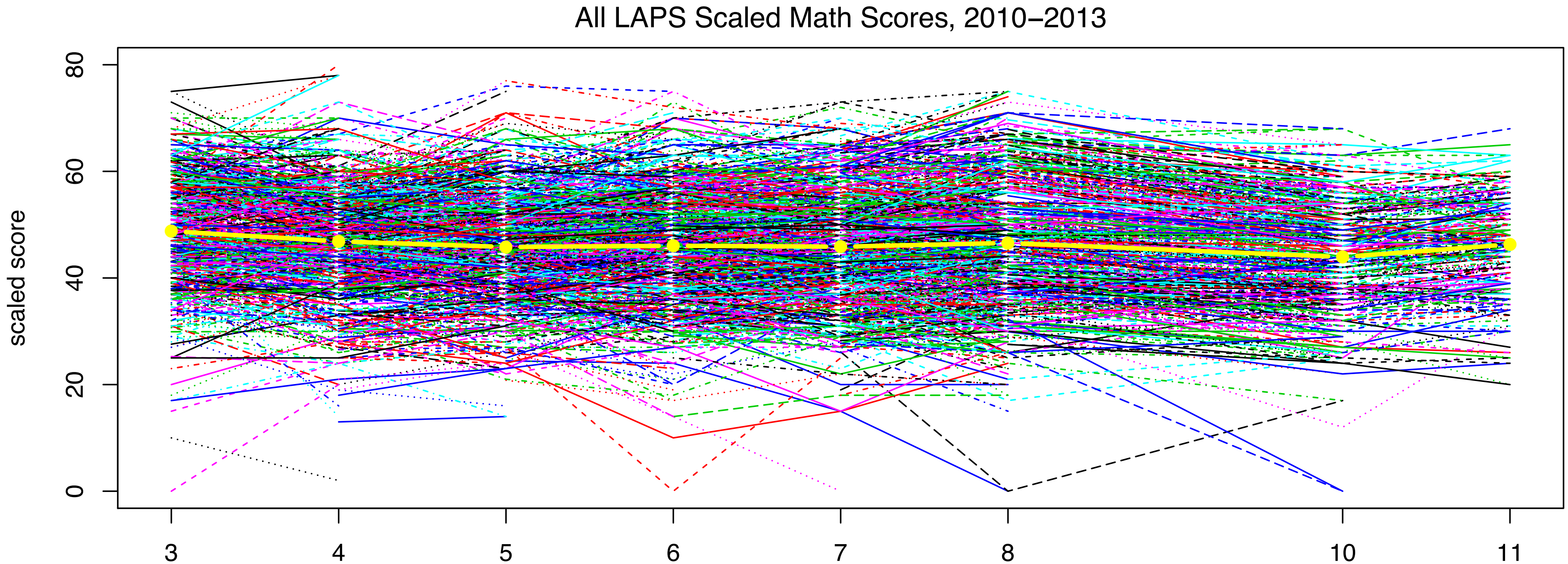
Test Scores & School Grades

Replaces "No Child Left Behind" scoring system

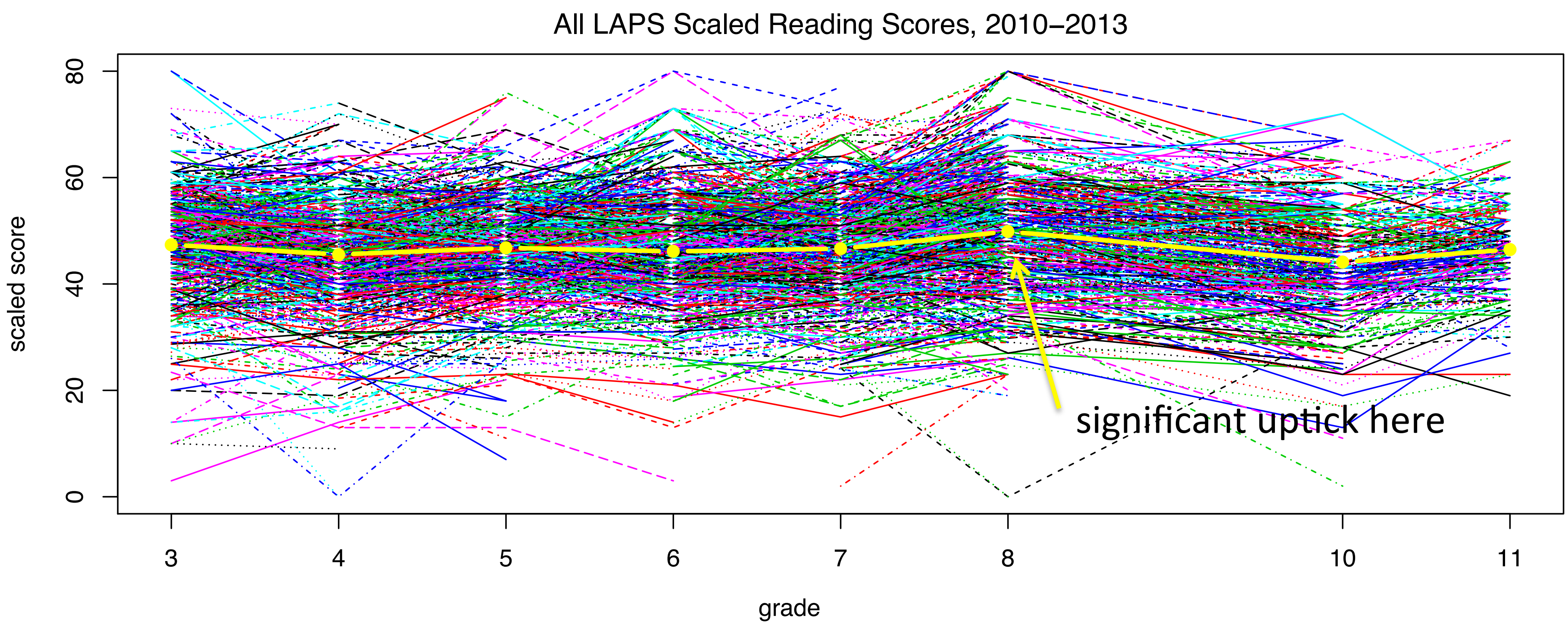
		Grade	School Points	Possible Points
Current Standing How did students perform in the most recent school year? Students are tested on how well they met targets for their grade level.	school score C grade 21.3	A	34.81	40
School Growth In the past 3 years did the school increase grade level performance? For example did this year's 3rd graders improve over last year's 3rd graders?	5.8	B	8.10	10
Student Growth of Highest Performing Students How well did the school help individual students improve? The highest performing students are those whose prior scores placed them in the top three quarters (75%) of their school. Individual student growth over the past 3 years is compared to the state benchmark.	7.2	B	9.70	20
Student Growth of Lowest Performing Students How well did the school help individual students improve? The lowest performing students are those whose prior scores placed them in the bottom quarter (25%) of their school. Individual student growth over the past 3 years is compared to the state benchmark.	15.3	F	1.47	20

Of all 100 points possible, 90 are based on these tests.
40 for performance + 50 for growth


All LAPS Scaled Math Scores, 2010–2013
scaled score
0
20
40
60
80
3
4
5
6
7
8
10
11
All LAPS Scaled Reading Scores, 2010–2013
scaled score
0
20
40
60
80
3
4
5
6
7
8
10
11
grade

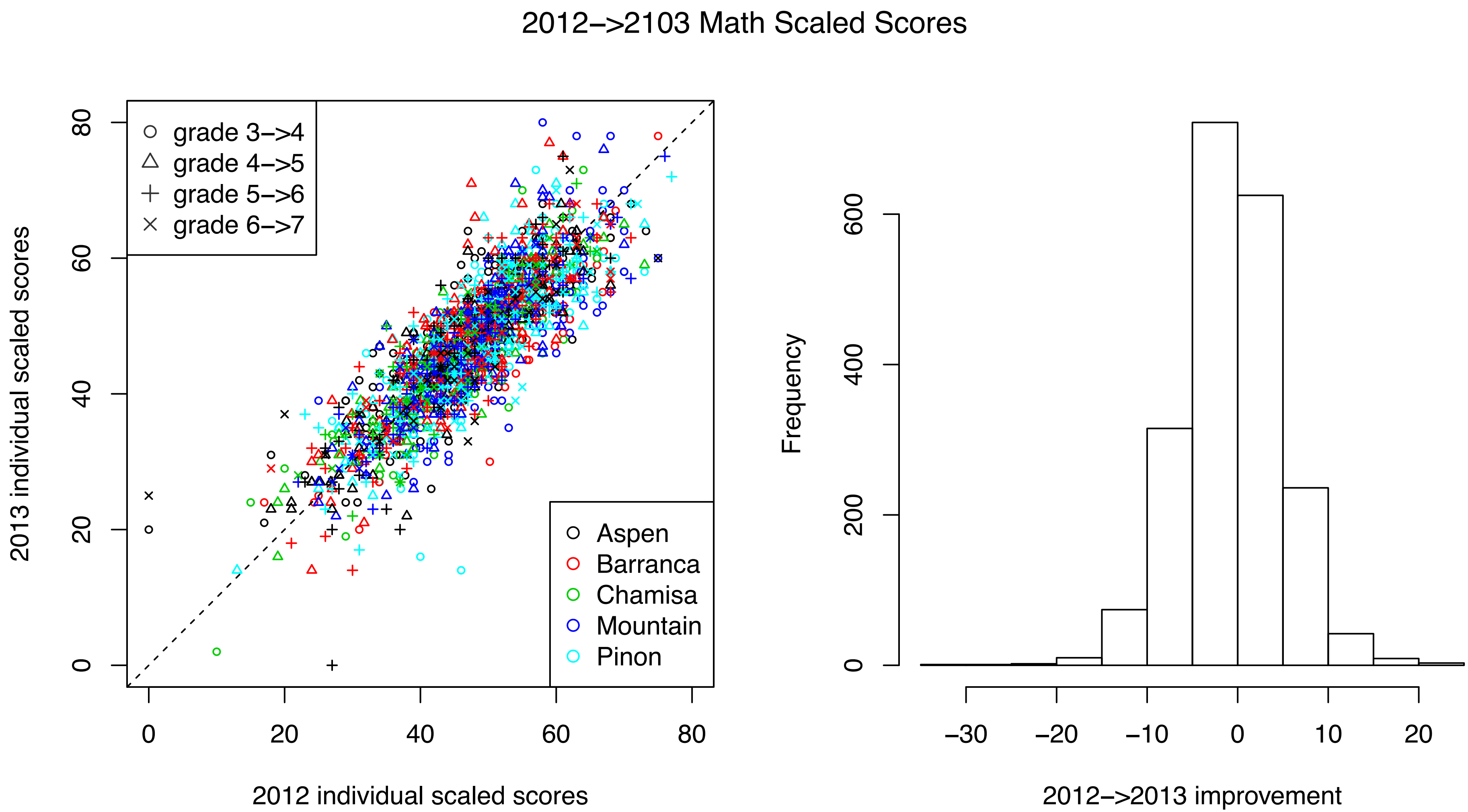


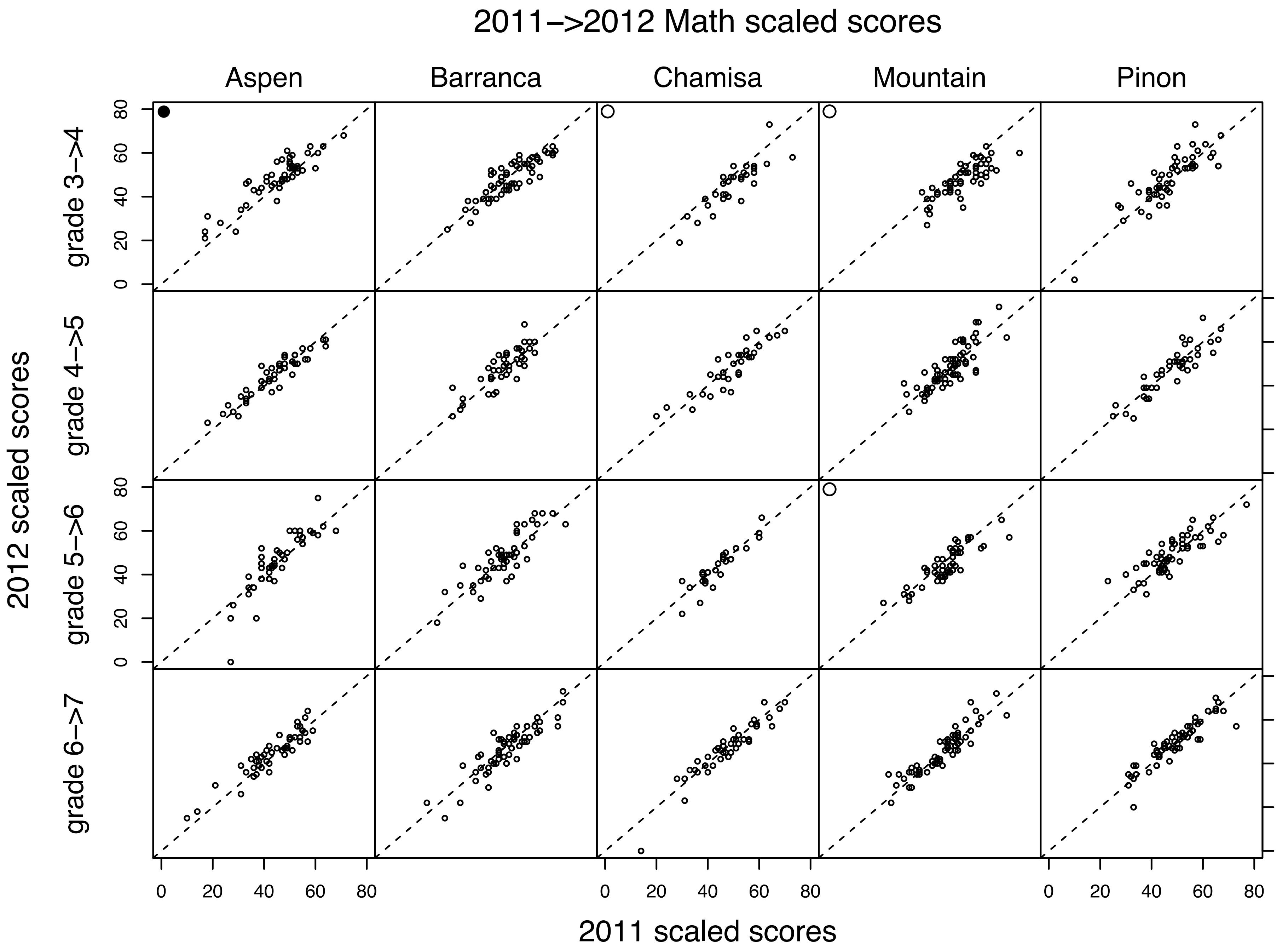
All LAPS Scaled Math Scores, 2010–2013
scaled score
0
20
40
60
80
3
4
5
6
7
8
10
11



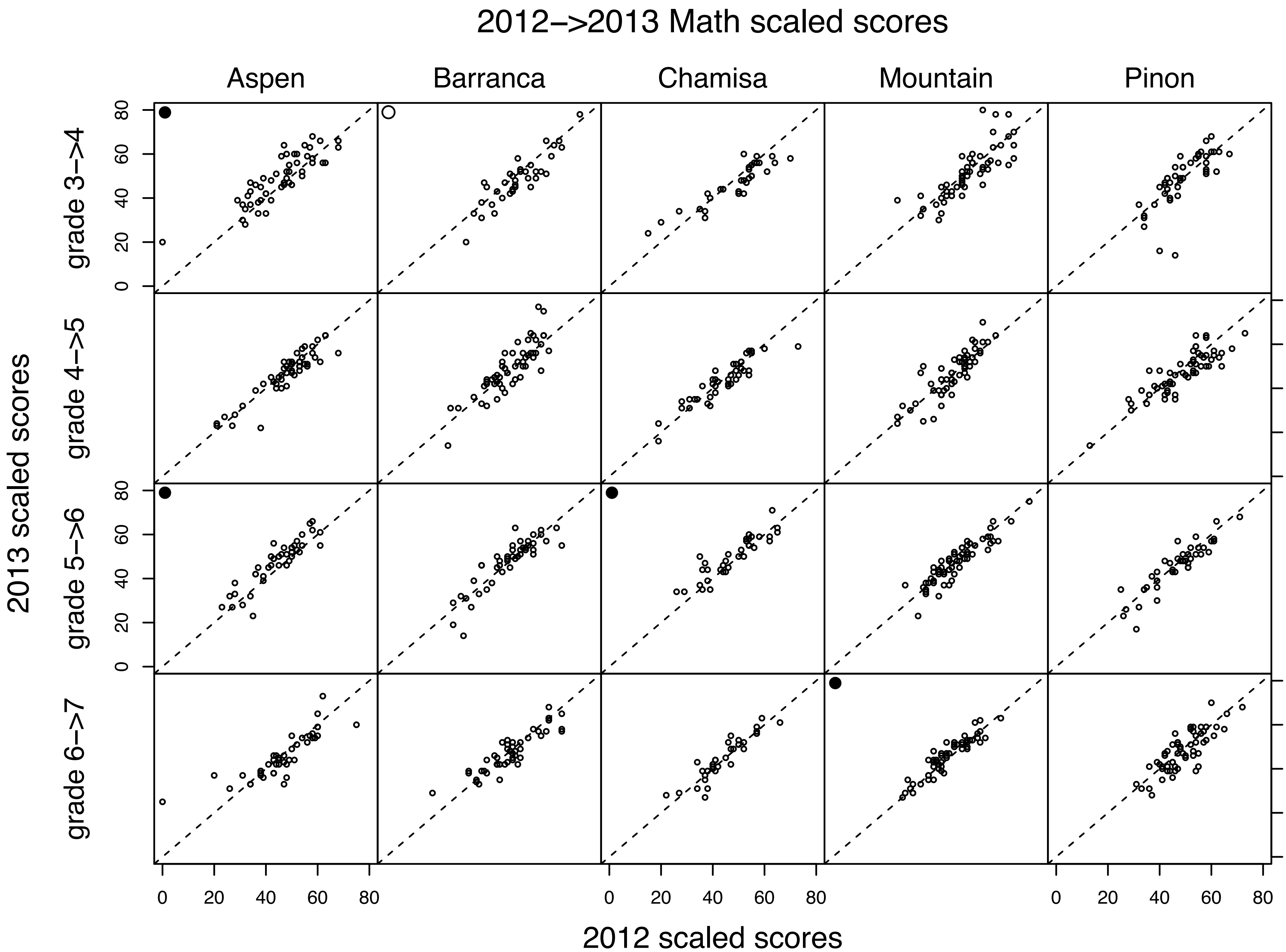
All LAPS Scaled Reading Scores, 2010–2013
scaled score
0
20
40
60
80
3
4
5
6
7
8
10
11
grade
significant uptick here

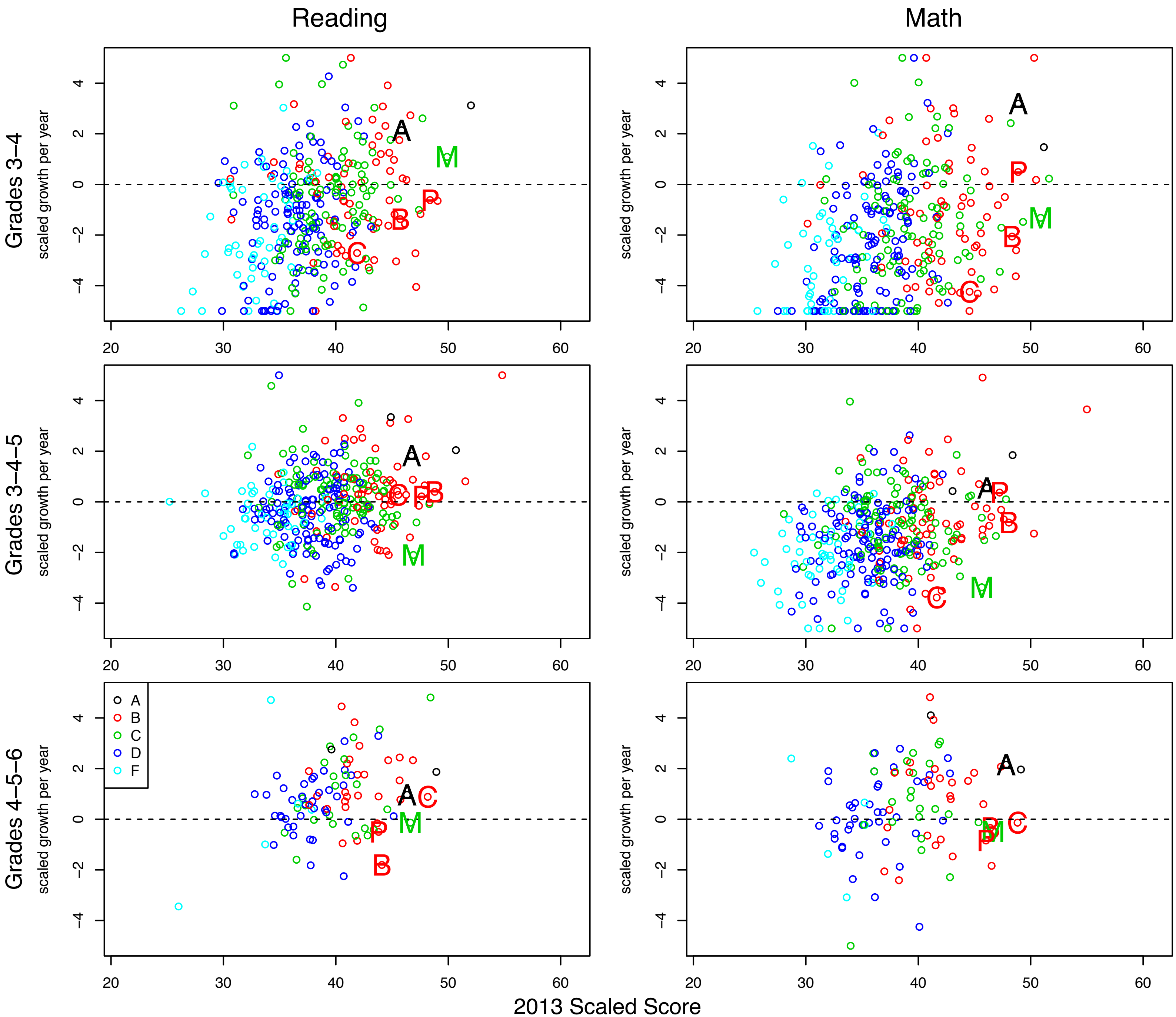
Math: (3 4 5 6) ➔ (4 5 6 7)



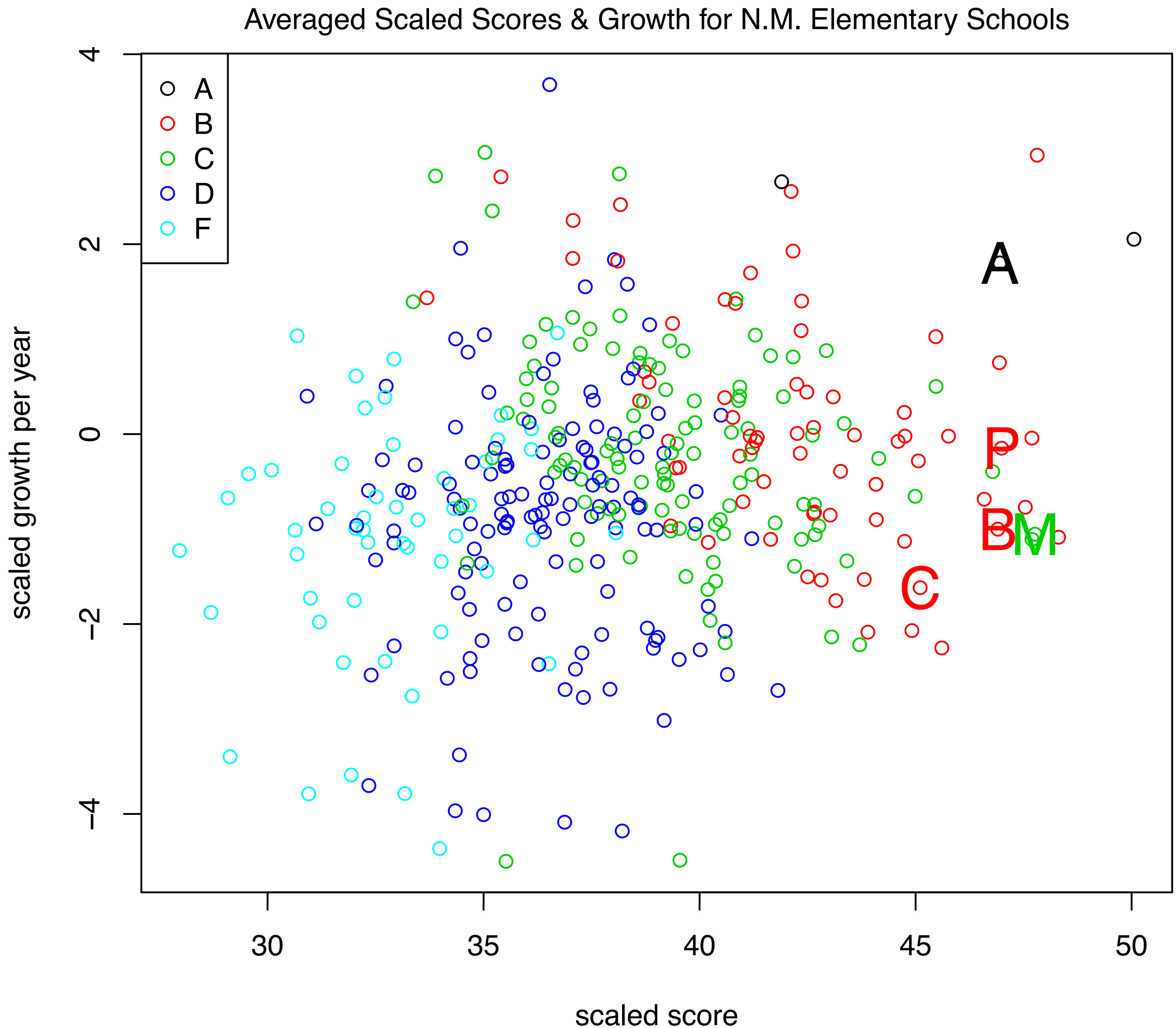


2011−>2012 Math scaled scores
Aspen
Barranca
Chamisa
Mountain
Pinon
grade 3−>4
grade 4−>5
grade 5−>6
grade 6−>7
2012 scaled scores
2011 scaled scores
0
20
40
60
80


2012−>2013 Math scaled scores
Aspen
Barranca
Chamisa
Mountain
Pinon
grade 3−>4
grade 4−>5
grade 5−>6
grade 6−>7
2013 scaled scores
2012 scaled scores
0
20
40
60
80

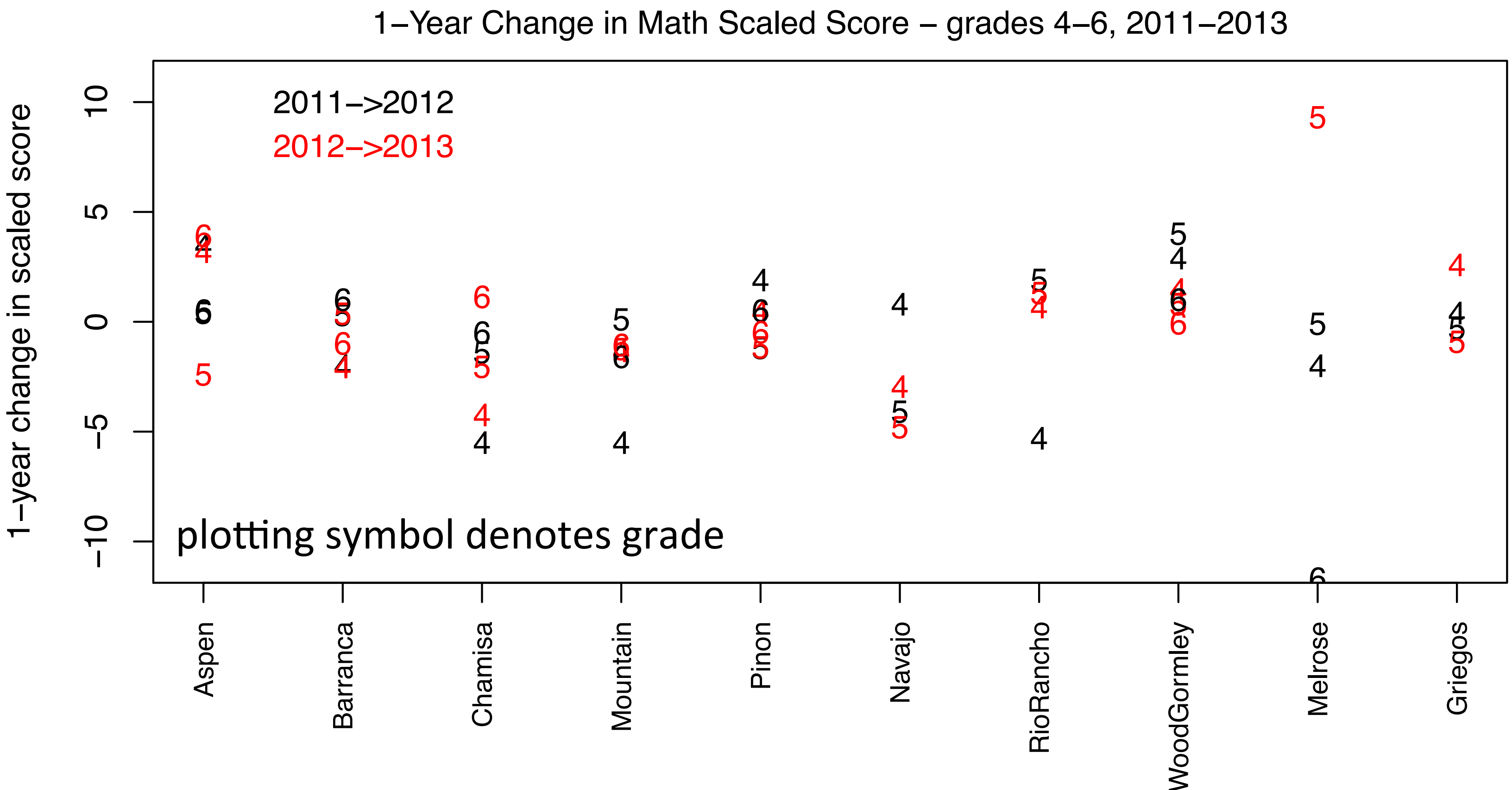


Reading
Math
Grades 3–4
Grades 3–4–5
Grades 4–5–6
scaled growth per year
2013 Scaled Score
A
B
C
D
F
M
P



Averaged Scaled Scores & Growth for N.M. Elementary Schools
A
B
C
D
F
A
P
B
M
C
scaled growth per year
scaled score
4
2
0
−2
−4
30
35
40
45
50

A simple ANOVA model



Model: growth = year*grade + school

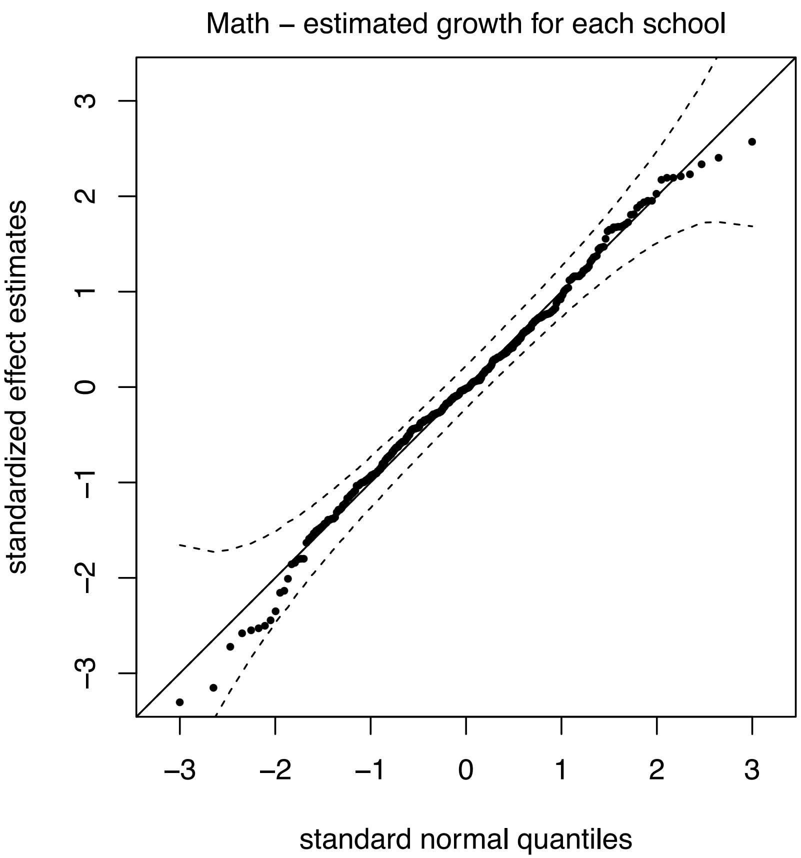
i={2011➔2012, 2012➔2013}, j={3➔4,4➔5,5➔6},

k=1,…,370 schools

$$y_{ijk} = \beta_0 + \beta_{ij} + \beta_k + \epsilon_{ijk},\ \epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$$

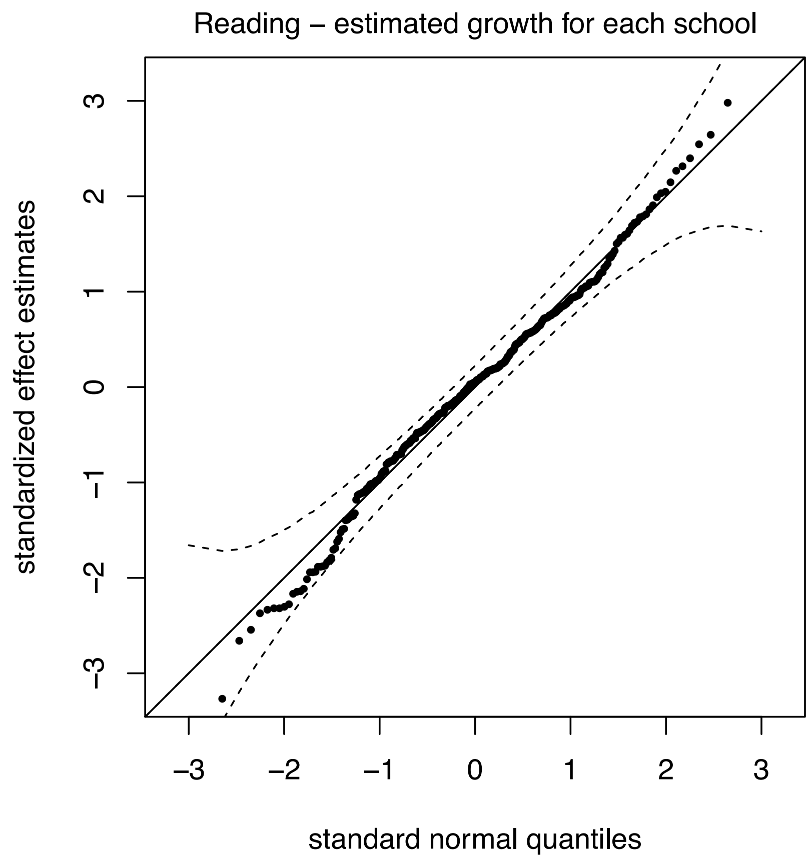
Analysis shows no significant school effect for growth

Estimated School Effects - Math



F=1.046 df=369,1235 p=.3

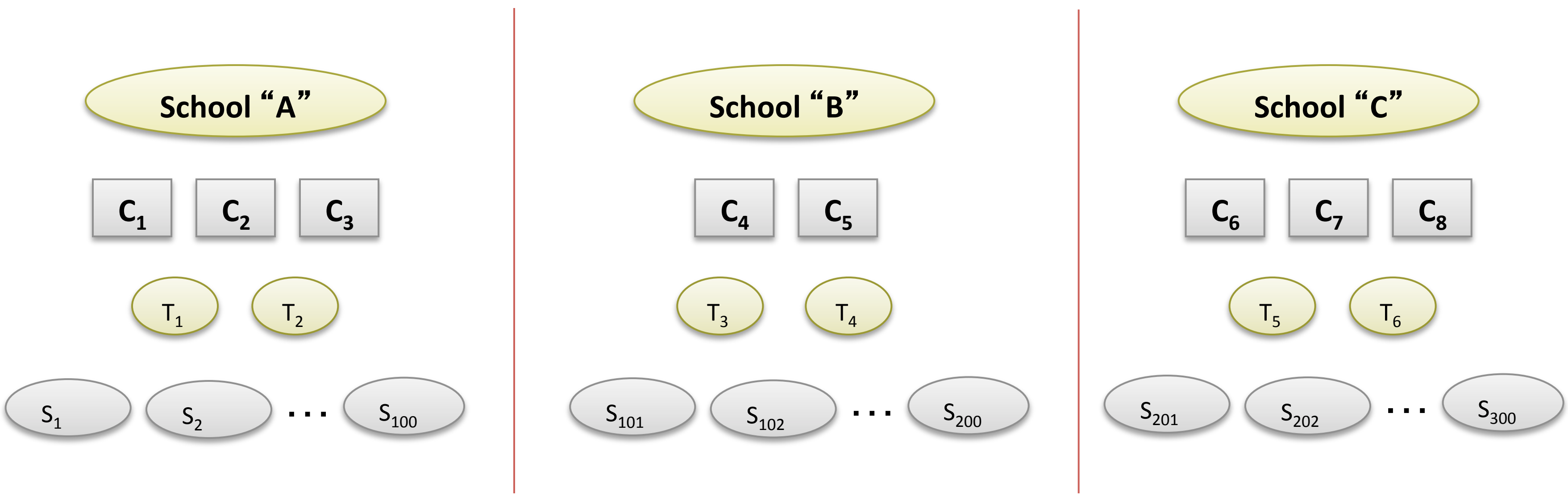
Estimated School Effects - Reading



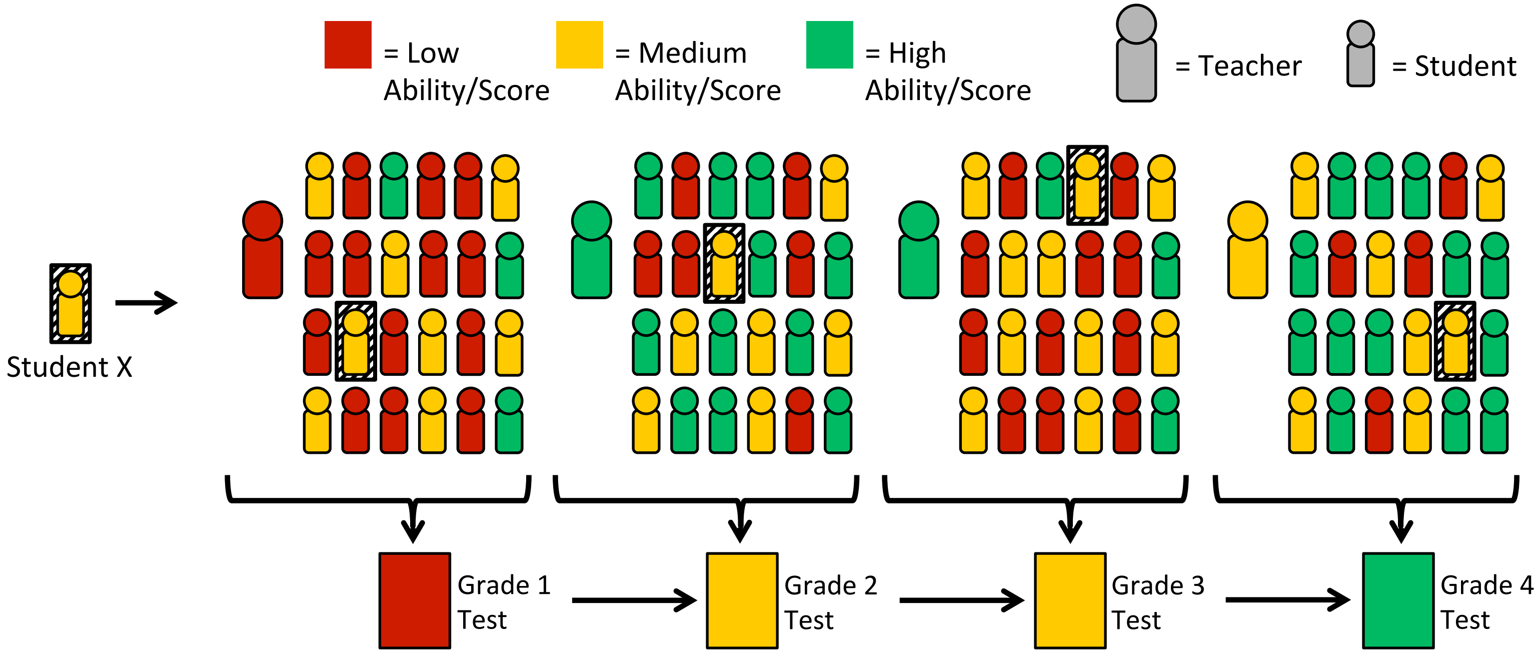
F=1.132 df=369,1235 p=.07

even less significant if no adjustment for year*grade is made

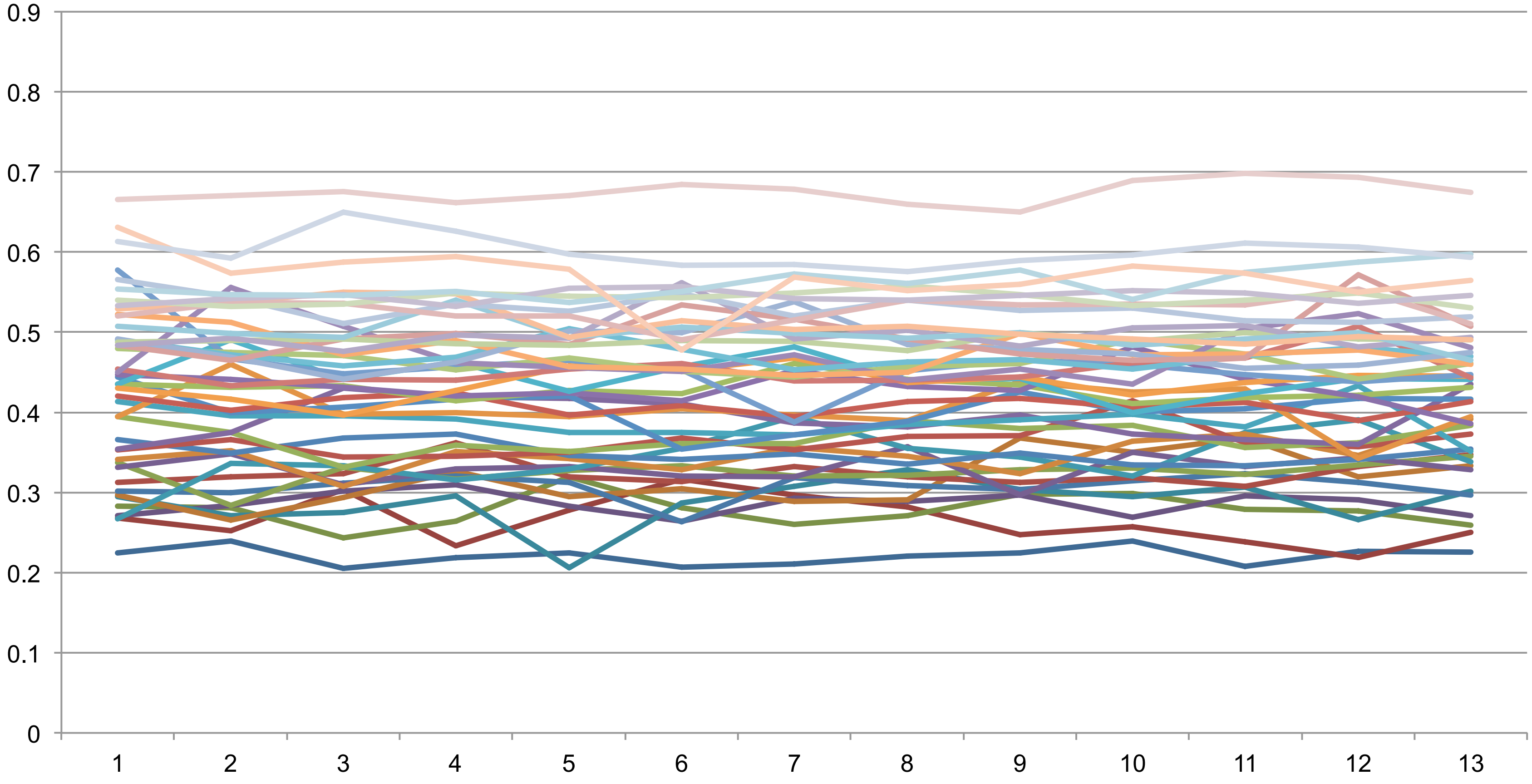
An Agent Based Simulation Model for a School System



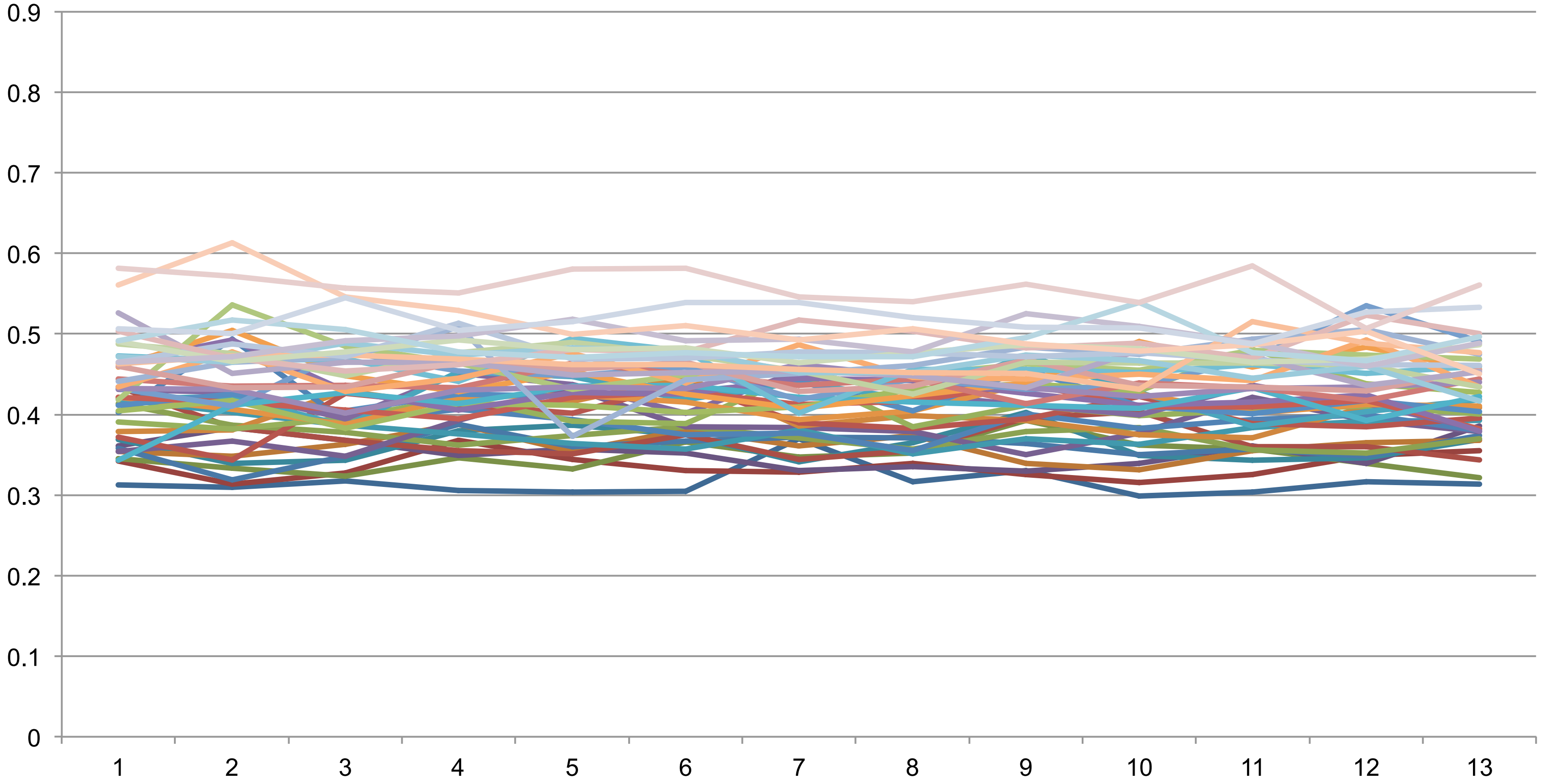
Visual Representation of Score Assignment



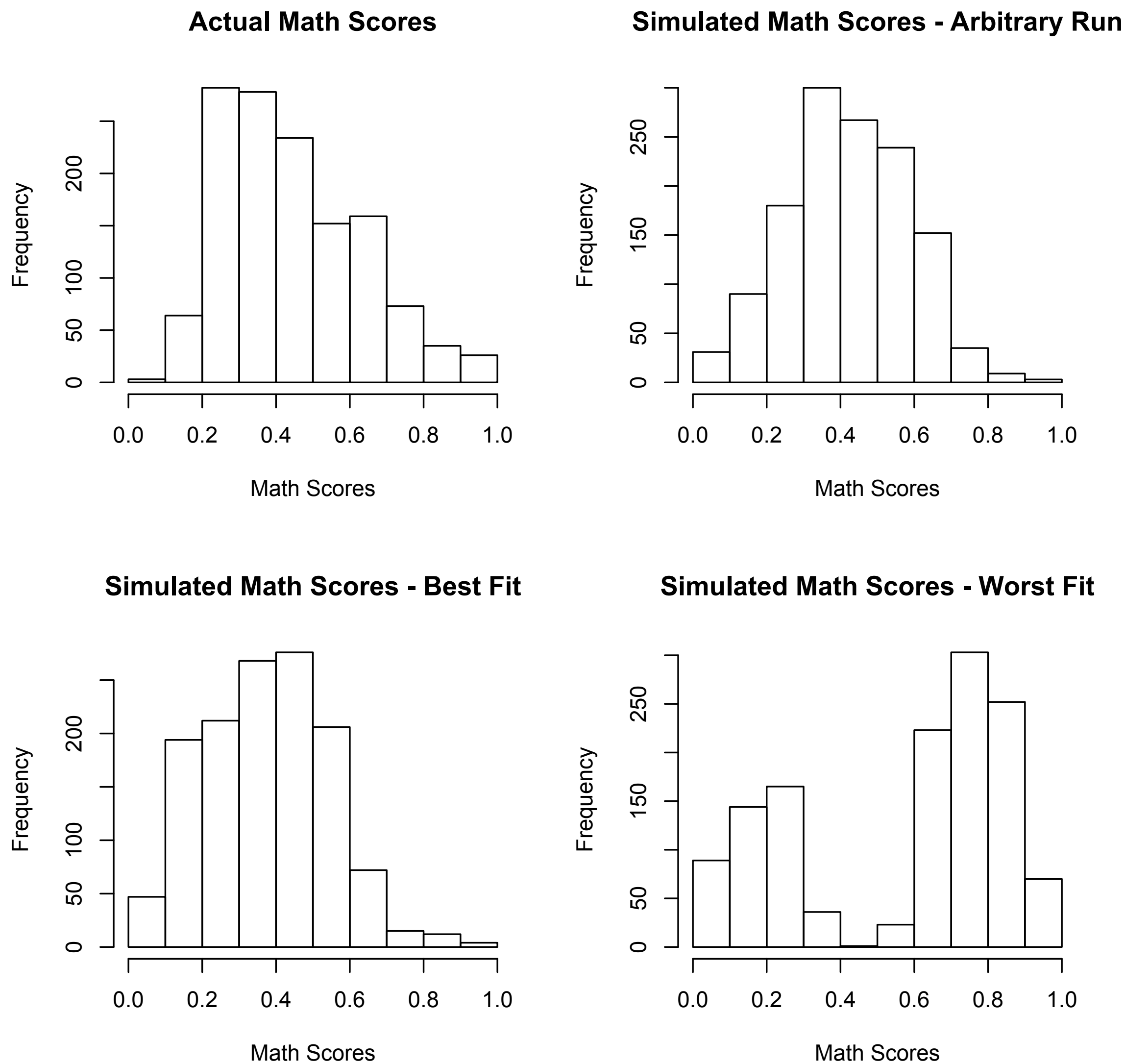
Sample score tracks, strict grouping by ability



Sample score tracks, completely random class assignment

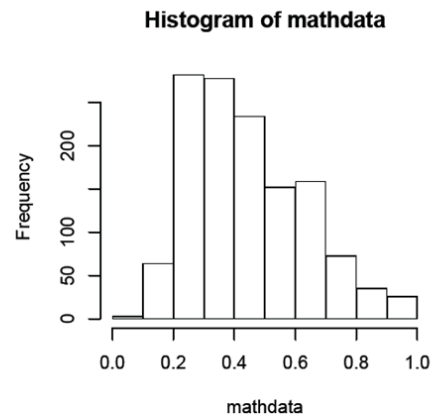


Actual and Simulated Data

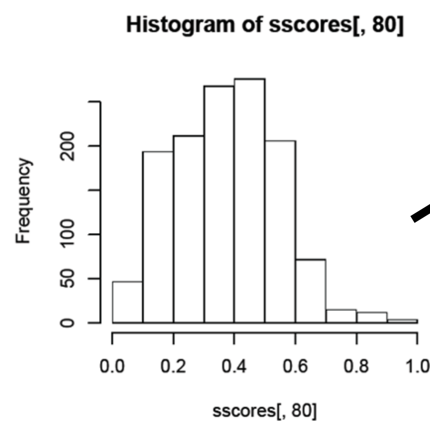


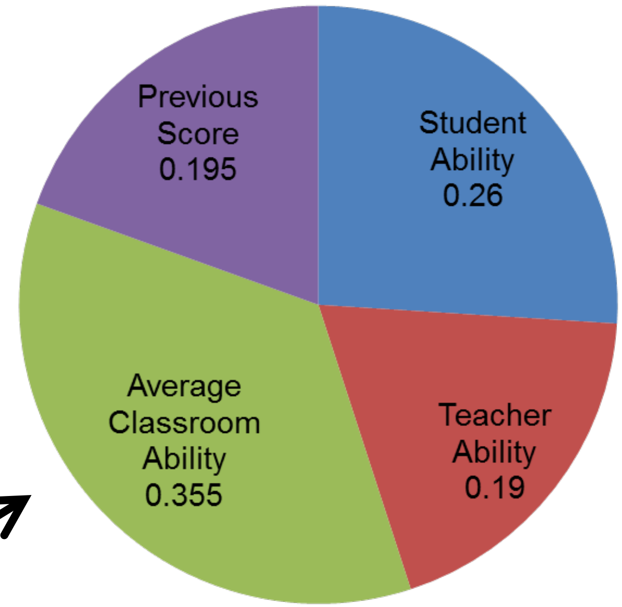
Calibration results

Actual 8th grade math scores



Simulated 8th grade math scores – closest match



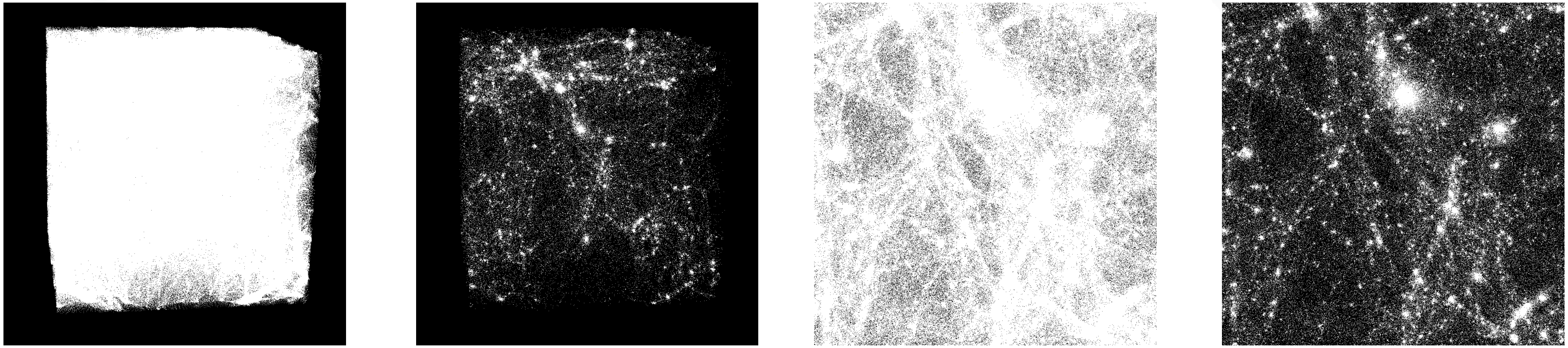


Weights resulting in closest match

Integrated Stat/CS Analysis And Visualization Research

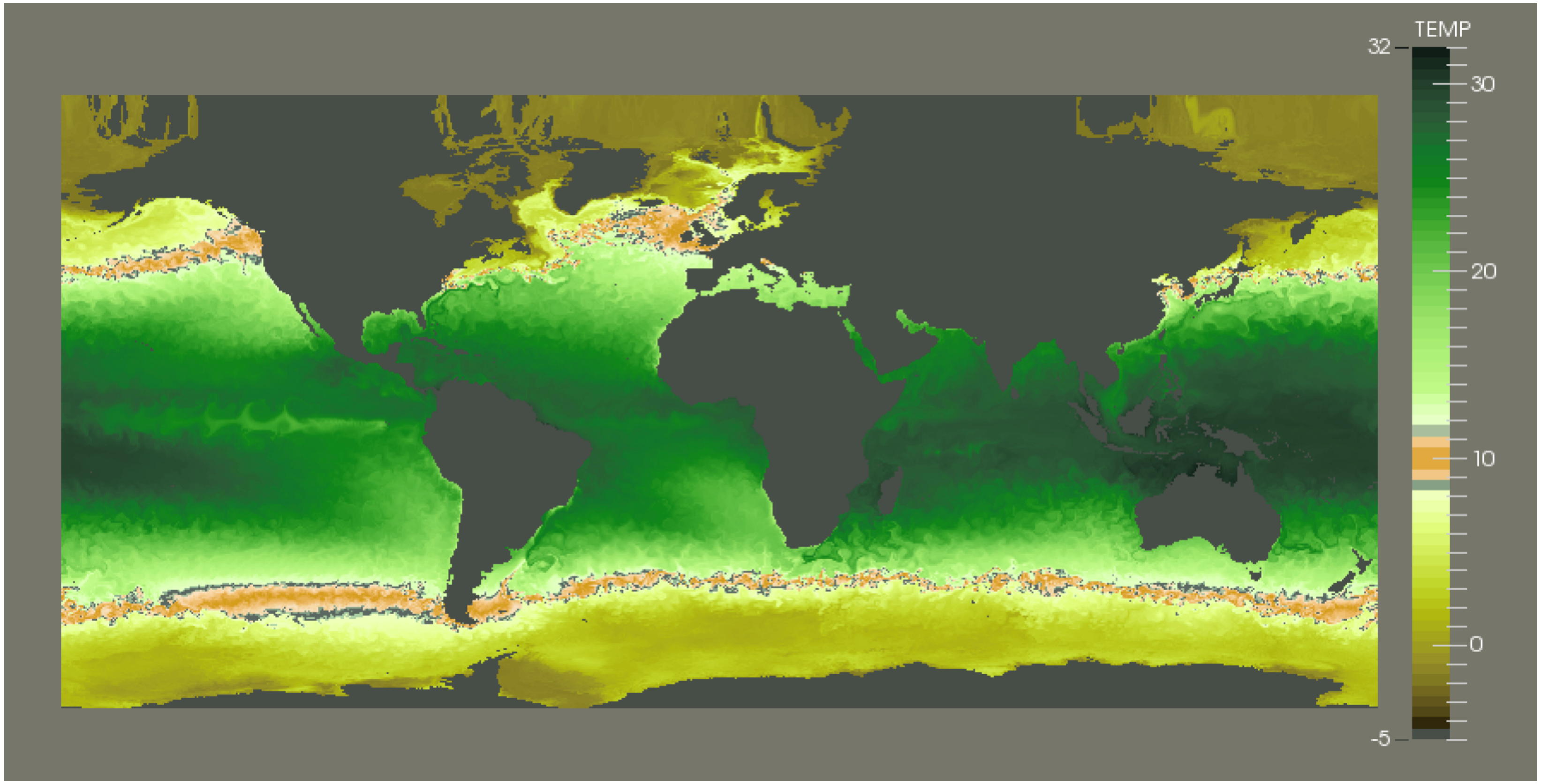
- In Situ Analysis
- Data Transformations
- Sampling of Large Data
- Color and Perception
- Uncertainty Regions for Multivariate Data
- Real-time Processing for Decision Support

Sampling:
Need data at an appropriate scale for specified analysis

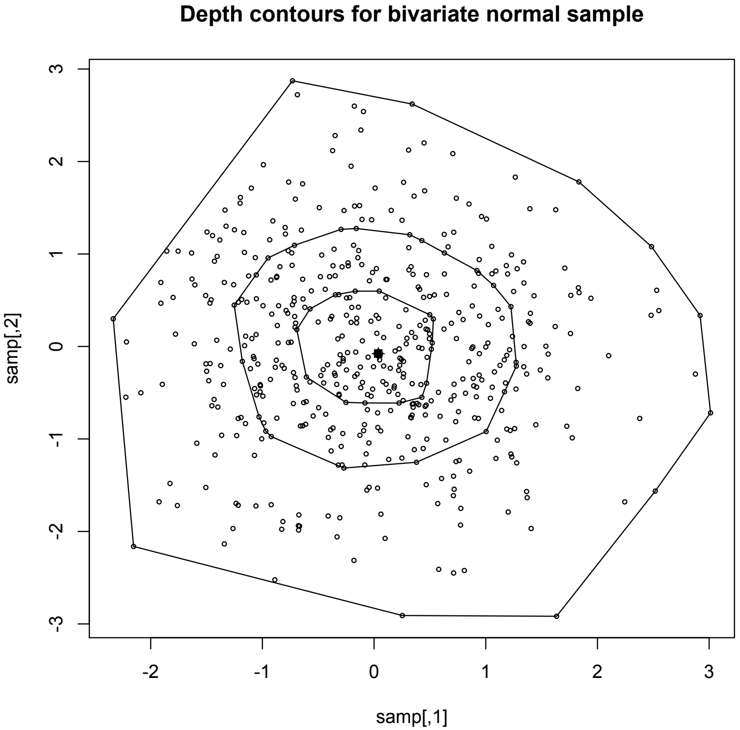


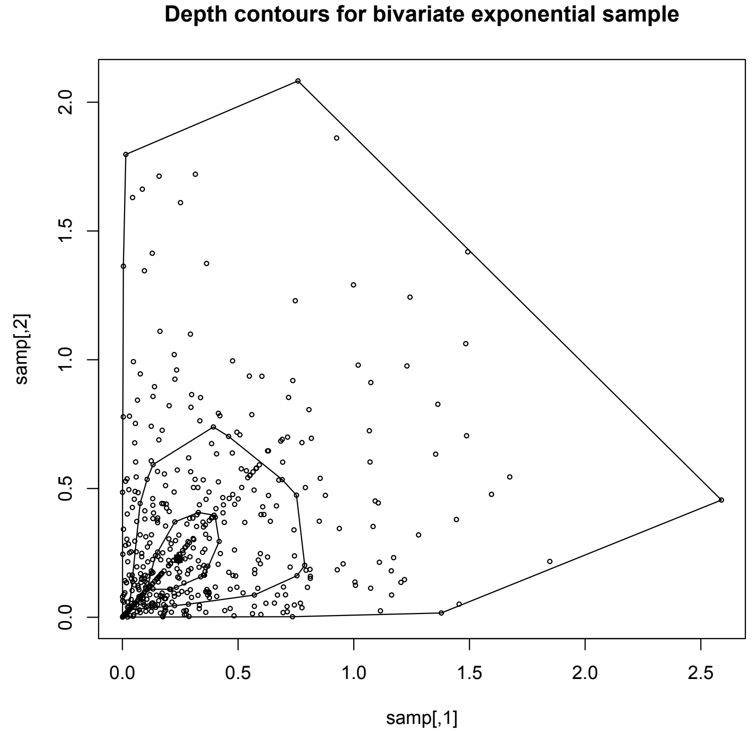
Particle data from a dark matter simulation at full resolution and samples generated via in-situ sampling.
From J. Woodring, J. Ahrens, J. Figg, J. Wendelberger, K. Heitmann (2011)

Color and Perception



Uncertainty Regions for Multivariate Data





THANK YOU!