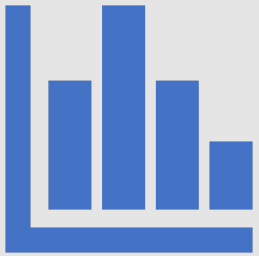
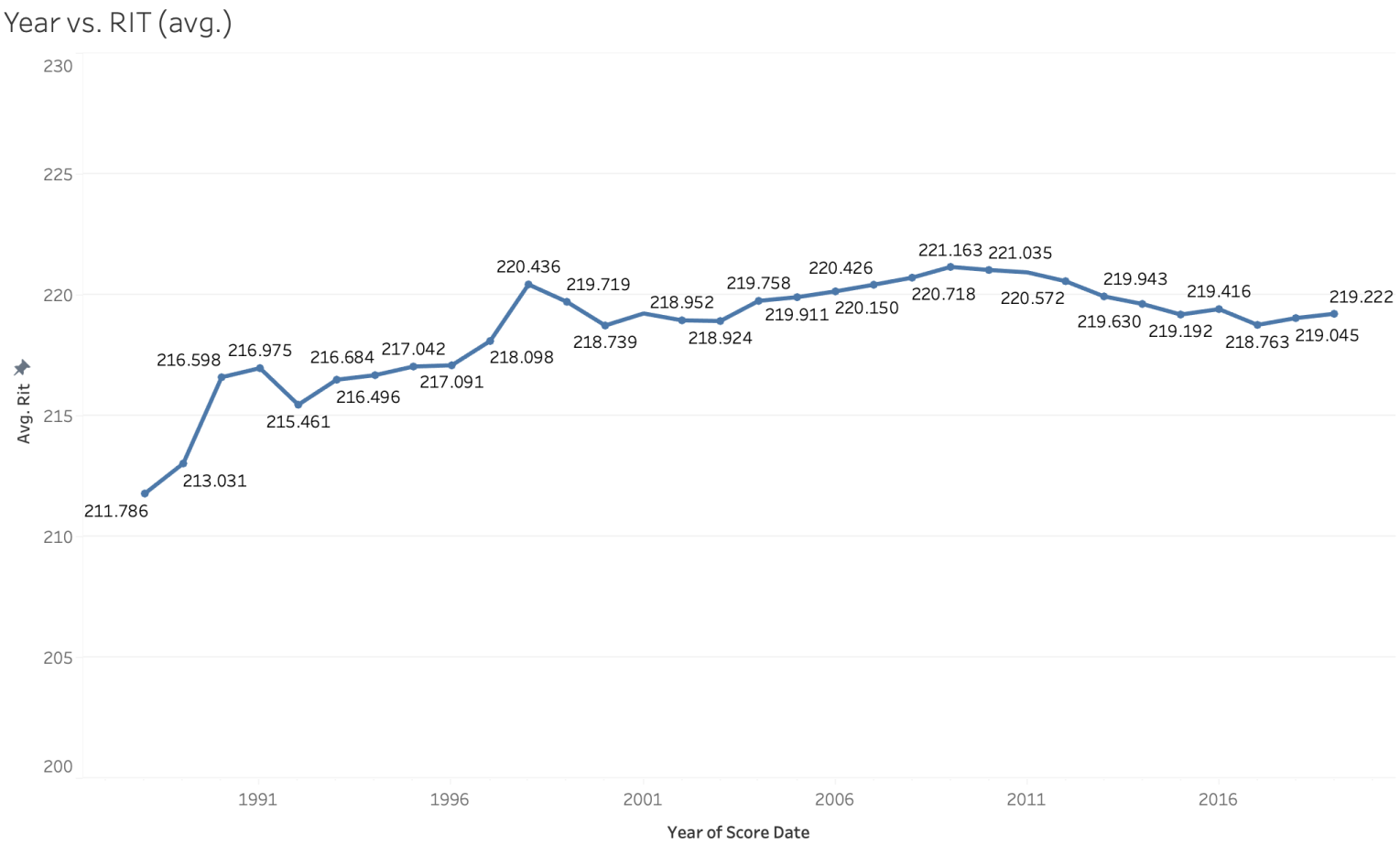
The Hewitt Foundation and the PASS Test

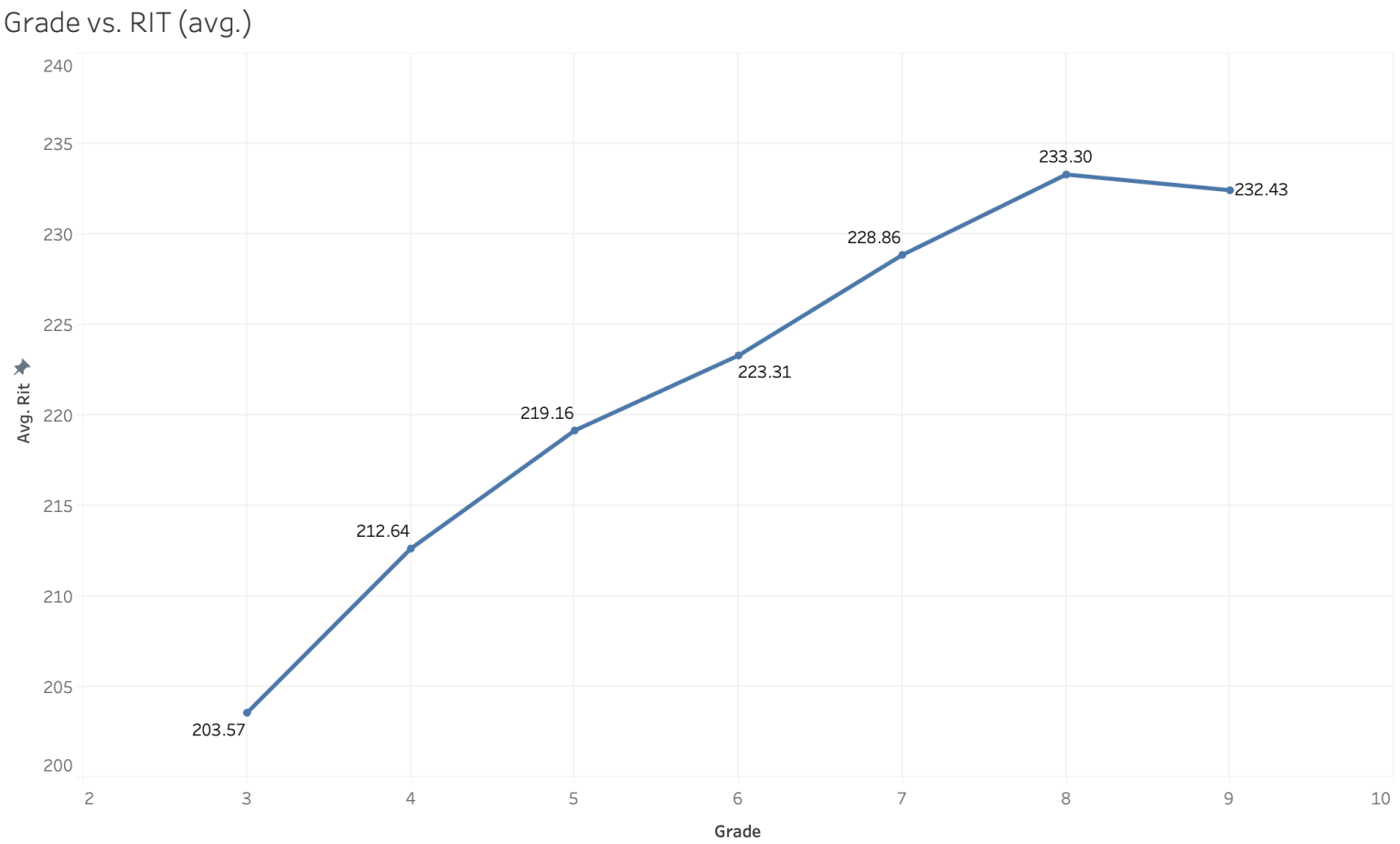
Graphical Observations /EDA



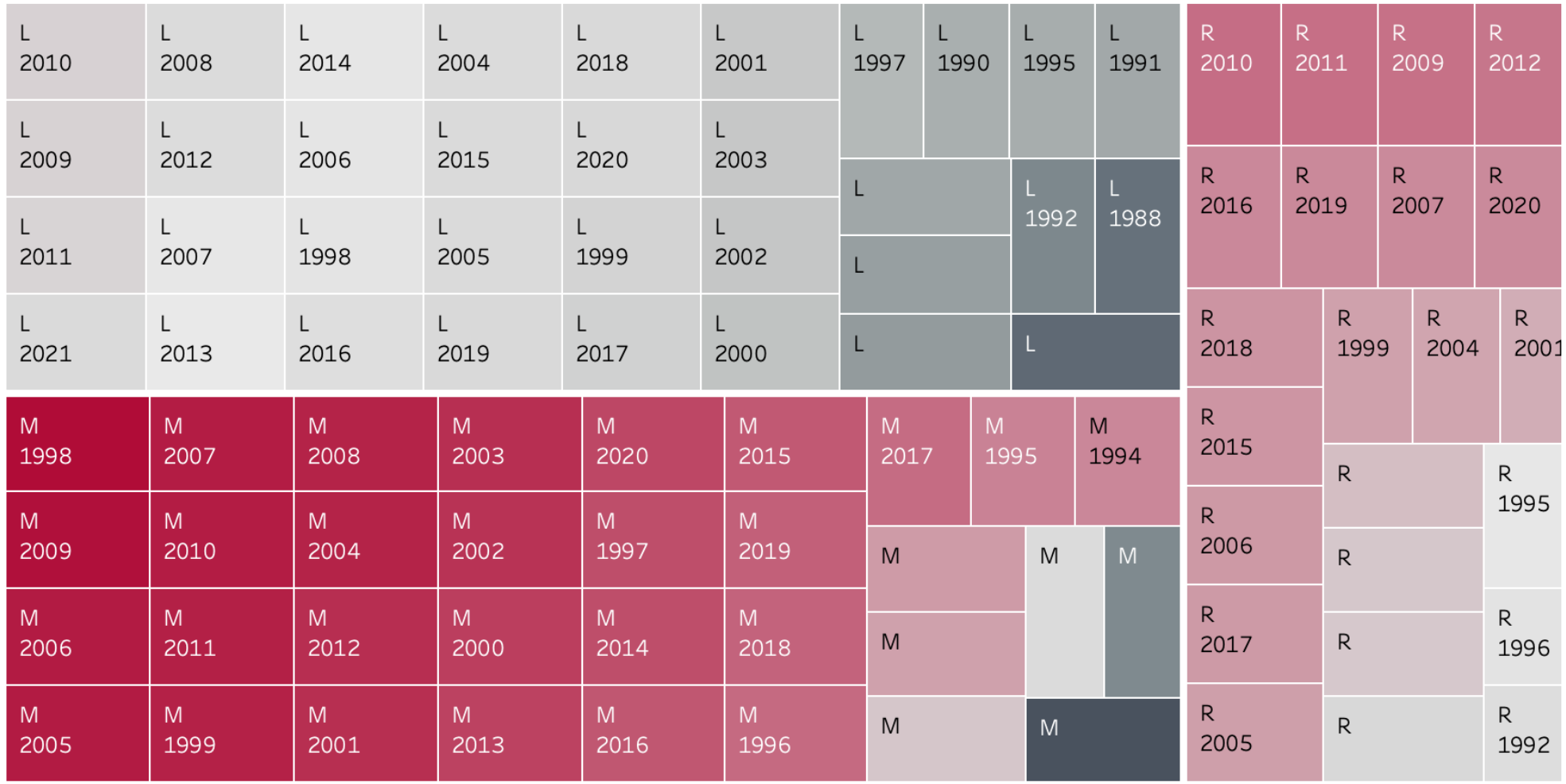
Plot of avg. RIT vs. Years



Plot of avg. RIT vs. Grades

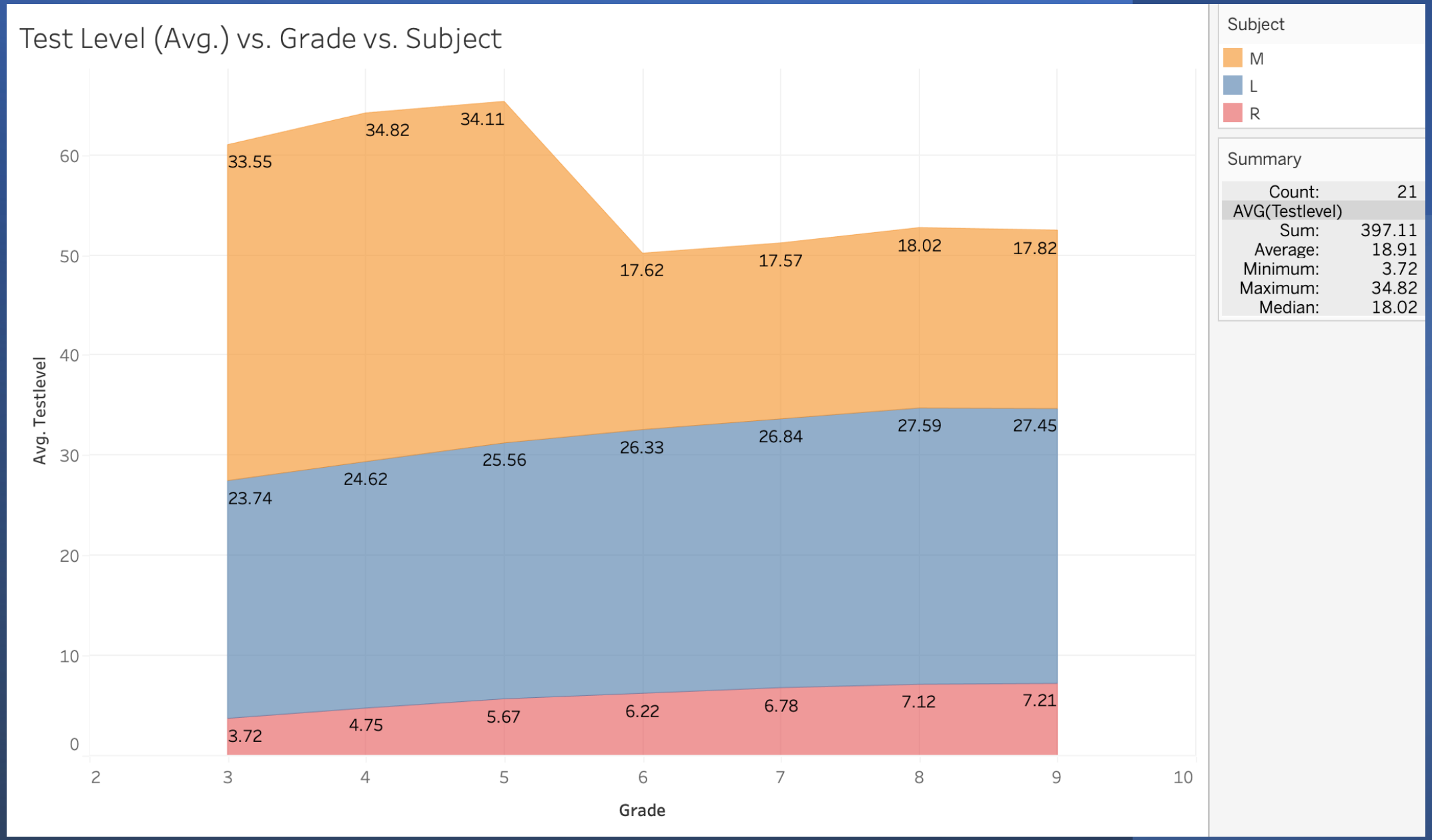


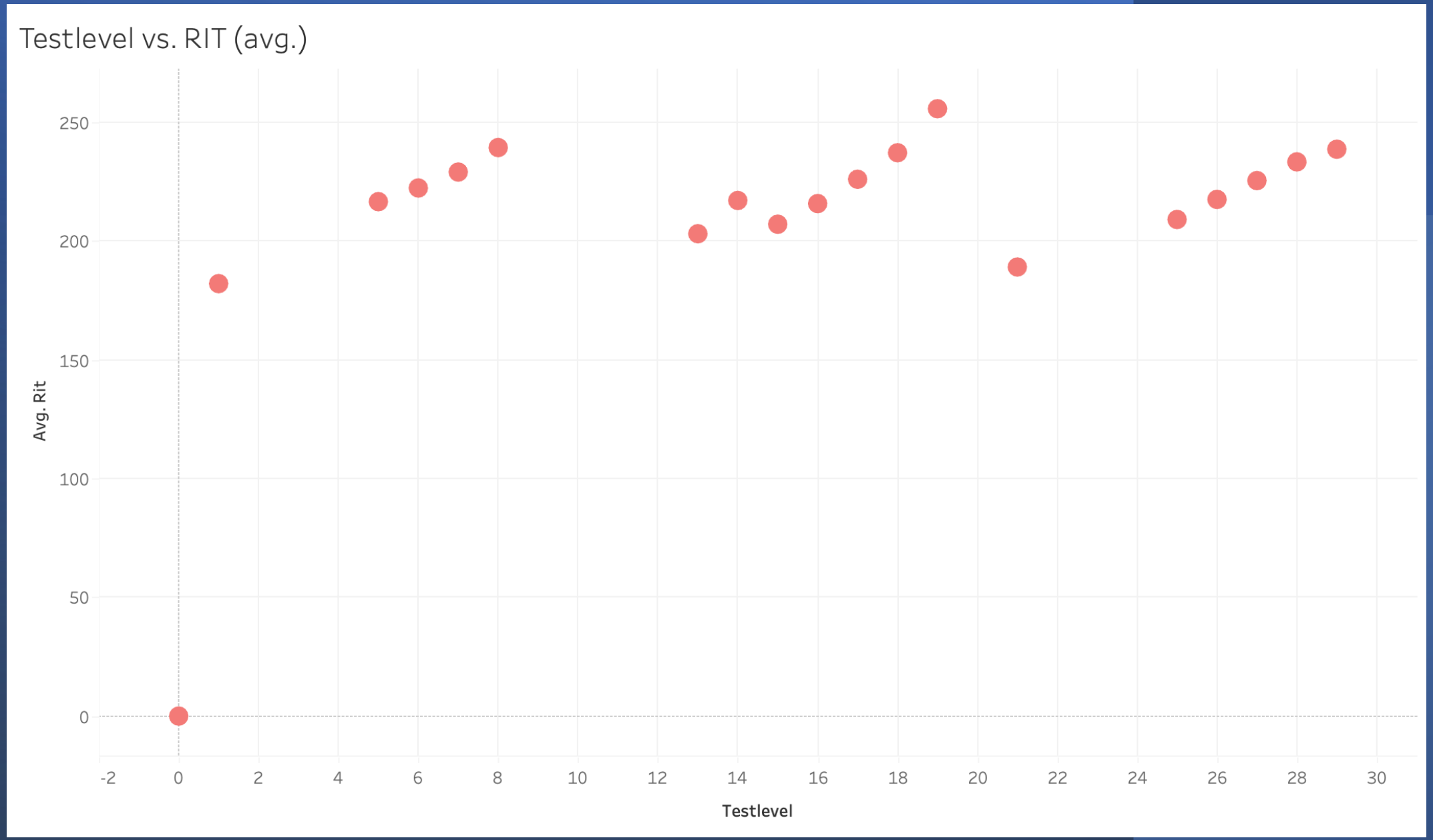
Subject vs. RIT (Avg) vs. Year



AVG(Rit)

209.73 225.09


Test Level (Avg.) vs. Grade vs. Subject
Subject
M
L
R
Summary
Count: 21
AVG(Testlevel)
Sum: 397.11
Average: 18.91
Minimum: 3.72
Maximum: 34.82
Median: 18.02
Avg. Testlevel
Grade
0
10
20
30
40
50
60
2
3
4
5
6
7
8
9
10
33.55
34.82
34.11
17.62
17.57
18.02
17.82
23.74
24.62
25.56
26.33
26.84
27.59
27.45
3.72
4.75
5.67
6.22
6.78
7.12
7.21

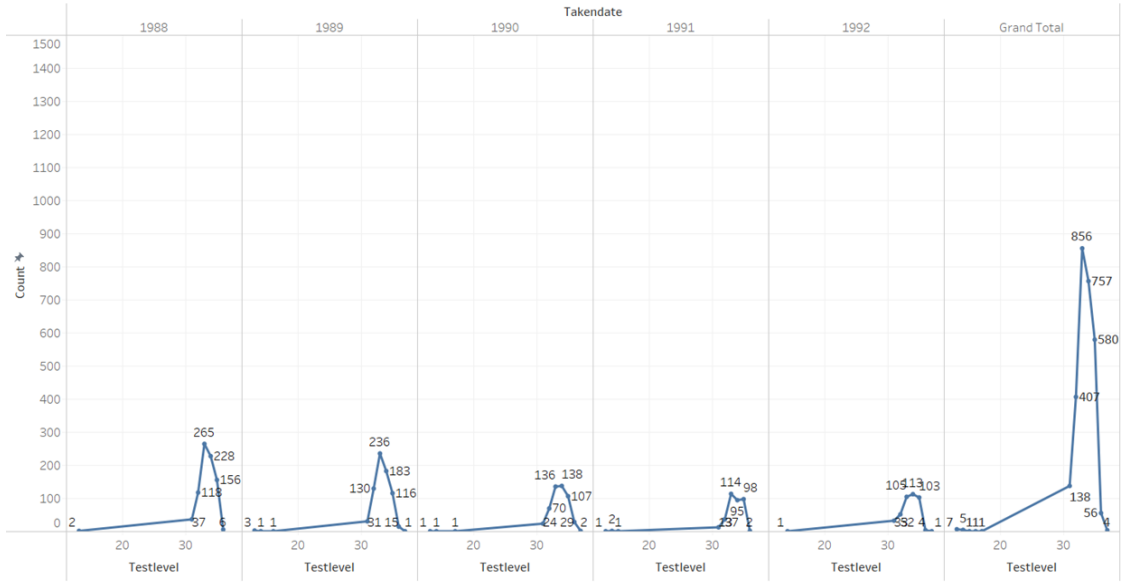

Testlevel vs. RIT (avg.)
250
200
150
100
50
0
Avg. Rit
-2
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Testlevel

Scatterplot for Count Vs TestLevel

MATH

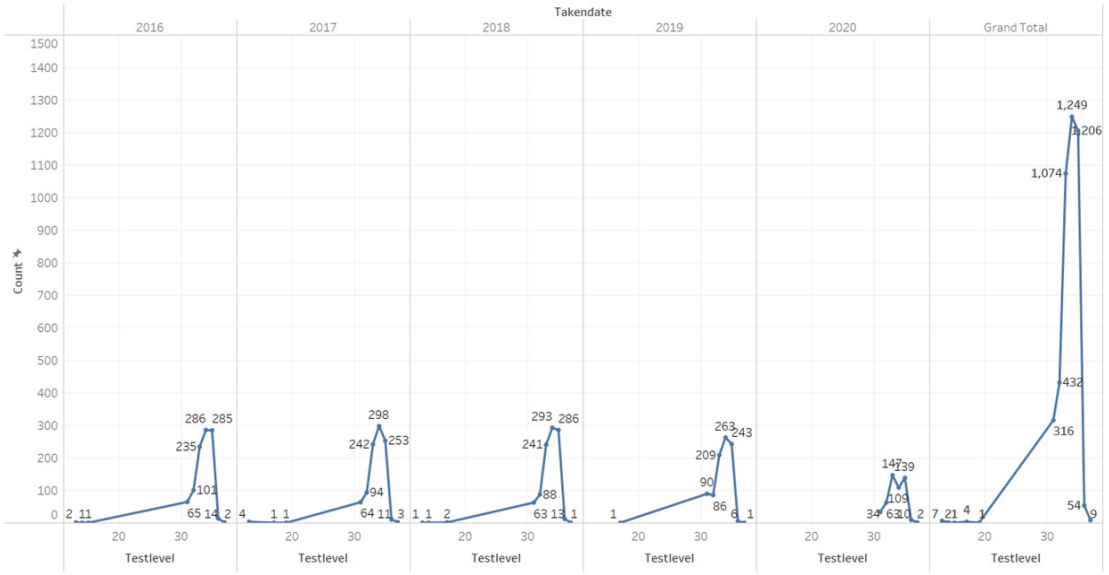
Count Vs TestLevel

Count Vs TestLevel **Grade 3 (1988-1992)**



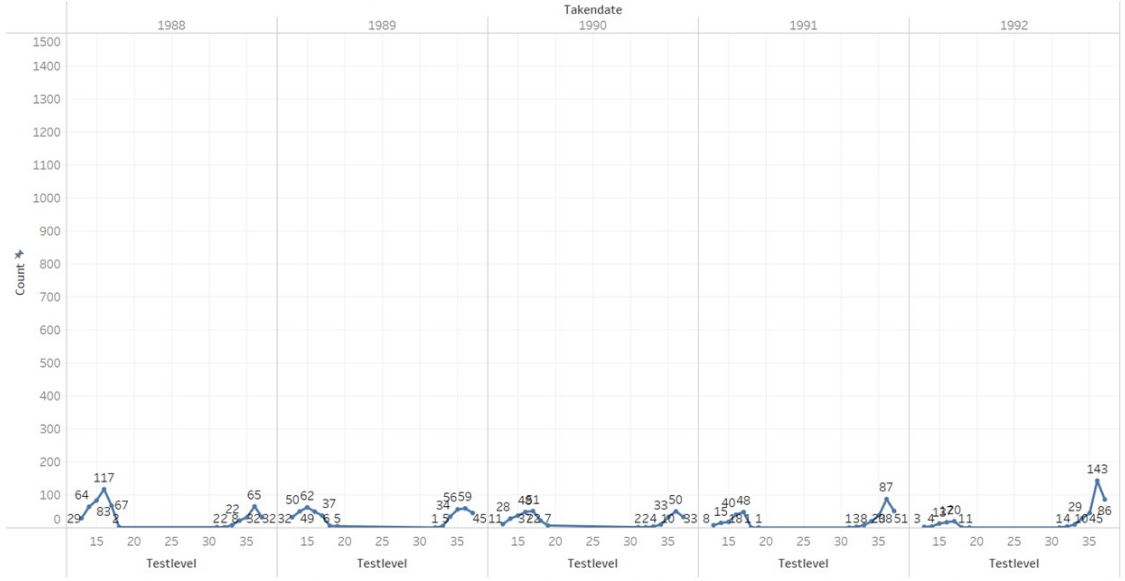
The trend of count of Math_all.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade and Testlevel. The Grade filter ranges from 3 to 3. The Testlevel filter excludes 0. The view is filtered on Takendate Year, which keeps 1988, 1989, 1990, 1991 and 1992.

Count Vs TestLevel **Grade 3 (2016-2020)**



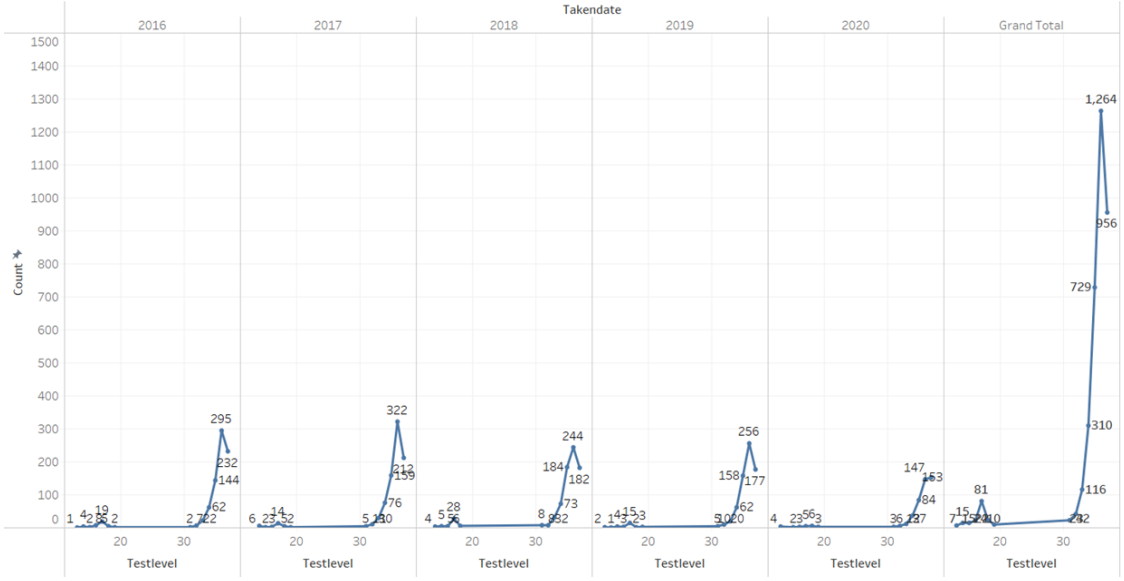
The trend of count of Math_all.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade and Testlevel. The Grade filter ranges from 3 to 3. The Testlevel filter excludes 0. The view is filtered on Takendate Year, which keeps 2016, 2017, 2018, 2019 and 2020.

Count Vs TestLevel **Grade 5 (1988-1992)**



The trend of count of Math_all.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade and Testlevel. The Grade filter ranges from 5 to 5. The Testlevel filter excludes 0. The view is filtered on Takendate Year, which keeps 1988, 1989, 1990, 1991 and 1992.

Count Vs TestLevel **Grade 5 (2016-2020)**

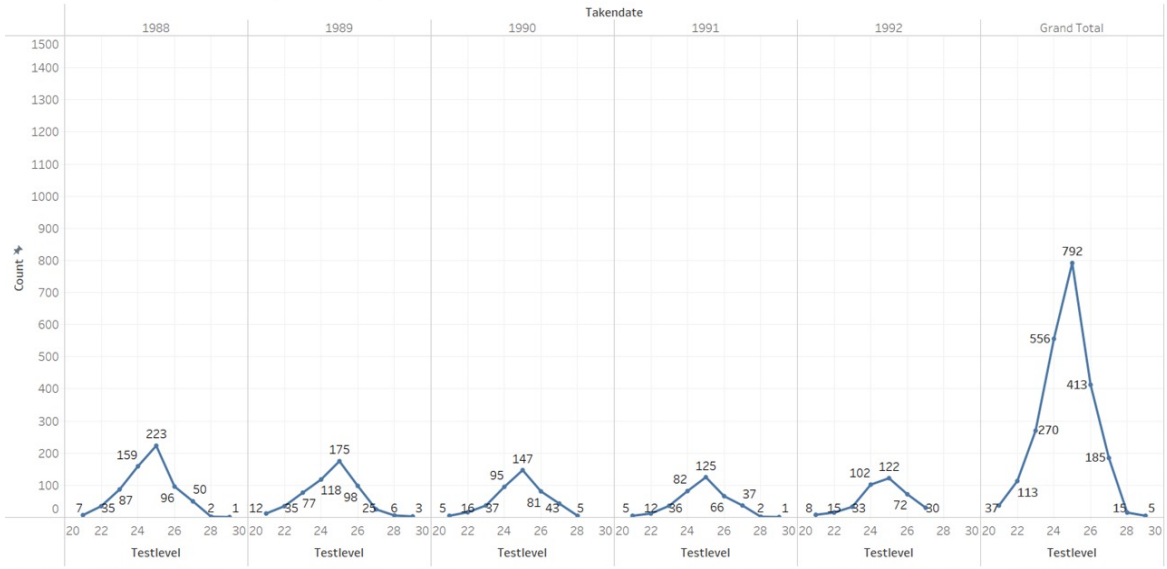


The trend of count of Math_all.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade and Testlevel. The Grade filter ranges from 5 to 5. The Testlevel filter excludes 0. The view is filtered on Takendate Year, which keeps 2016, 2017, 2018, 2019 and 2020.

Language & Arts

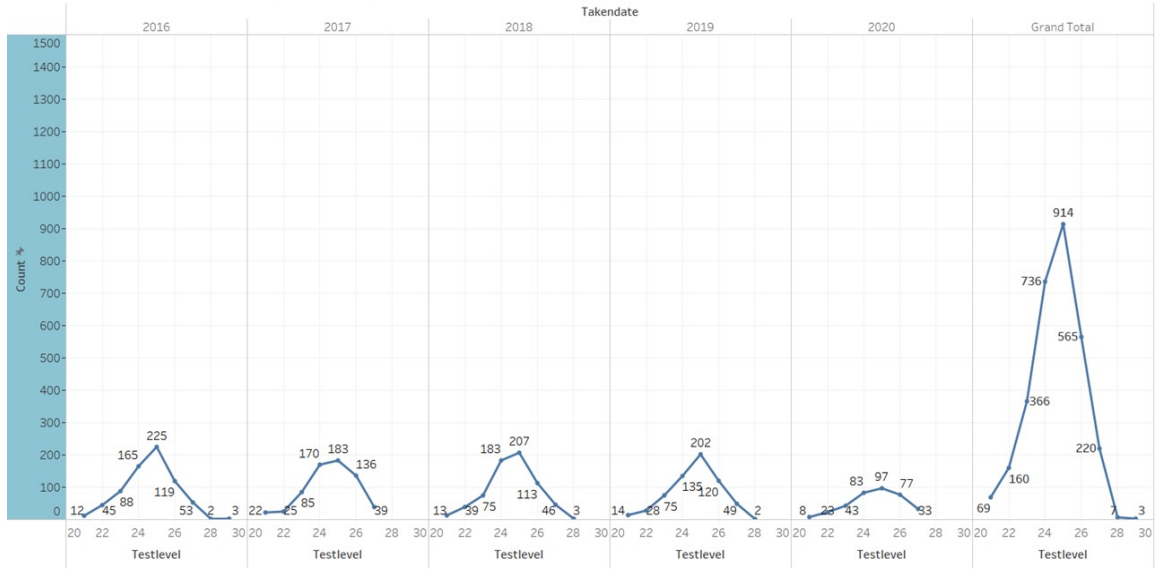
Count Vs TestLevel

Count Vs TakenDate **Grade 4 (1988-1992)**



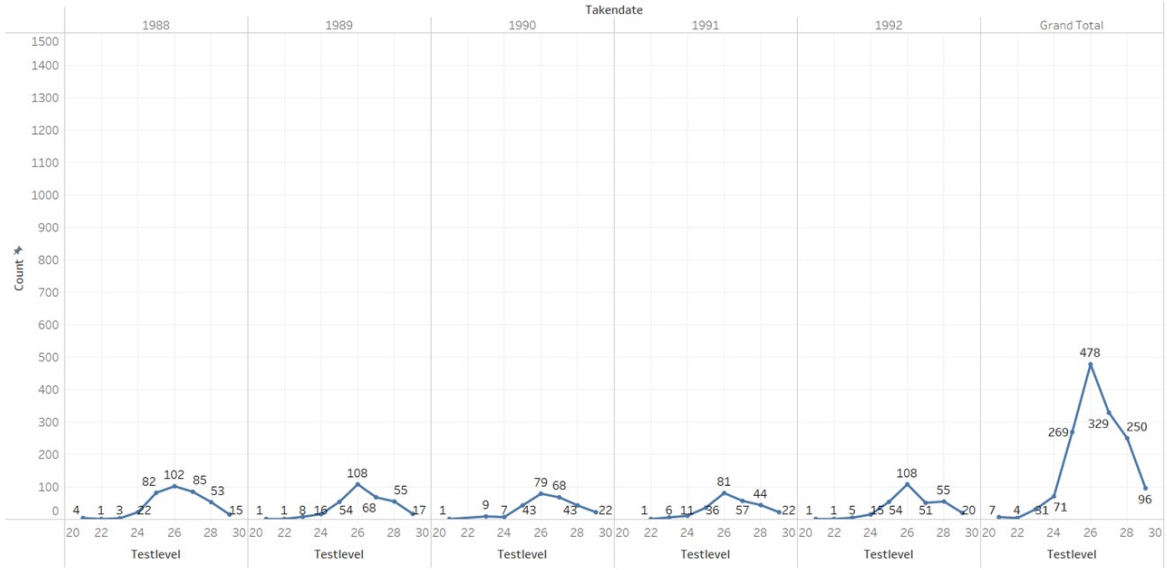
The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 4 to 4. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 1988, 1989, 1990, 1991 and 1992. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 394 members.

Count Vs TakenDate **Grade 4 (2016-2020)**



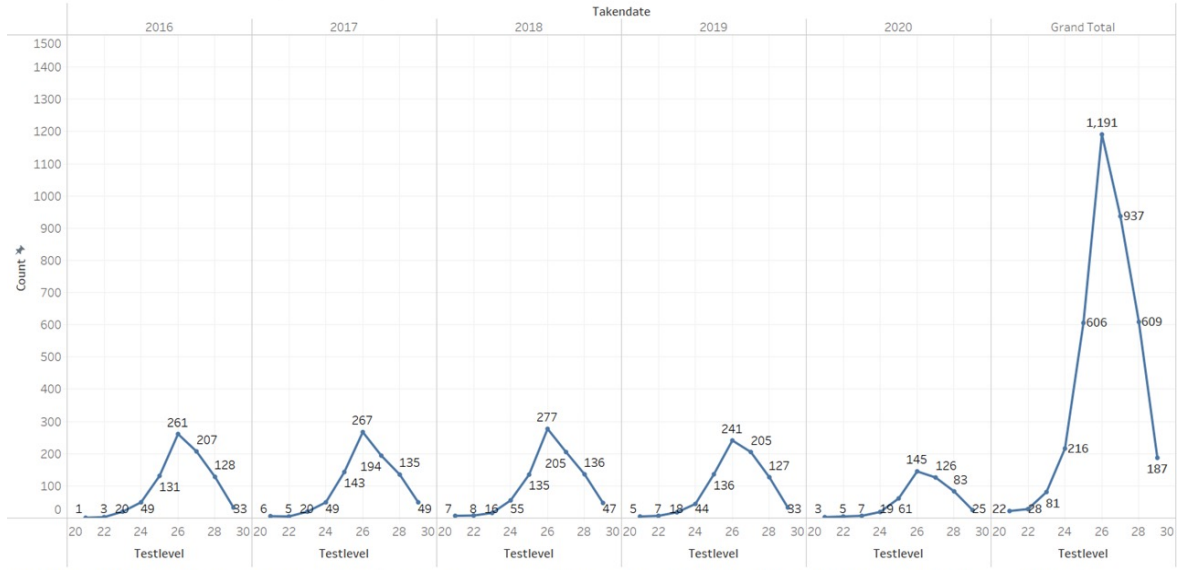
The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 4 to 4. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 2016, 2017, 2018, 2019 and 2020. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 392 members.

Count Vs TakenDate **Grade 6 (1988-1992)**



The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 6 to 6. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 1988, 1989, 1990, 1991 and 1992. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 393 members.

Count Vs TakenDate **Grade 6 (2016-2020)**

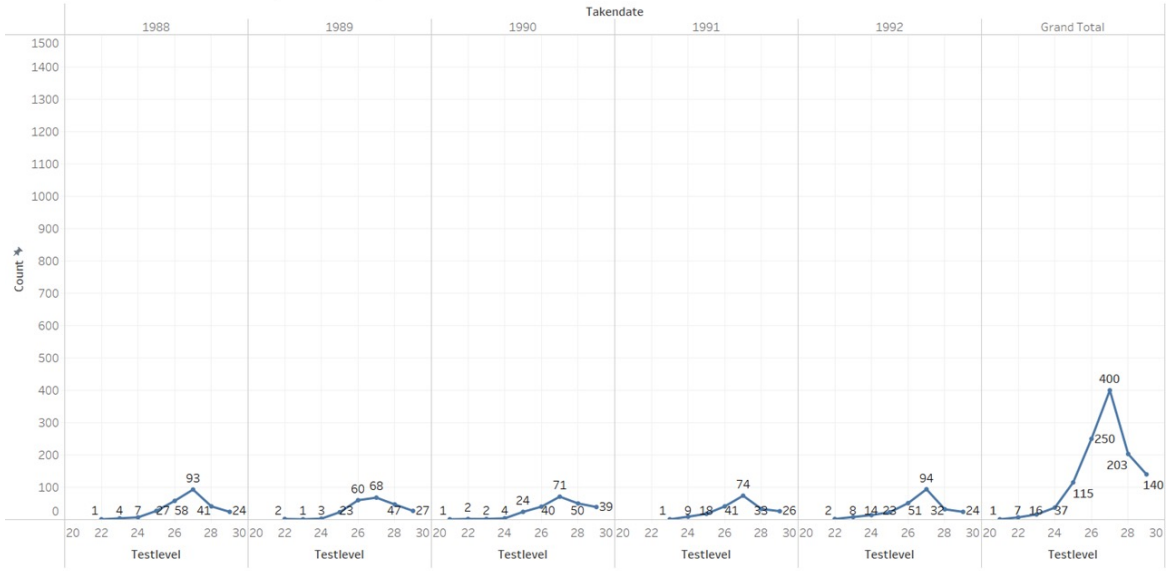


The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 6 to 6. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 2016, 2017, 2018, 2019 and 2020. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 394 members.

Reading

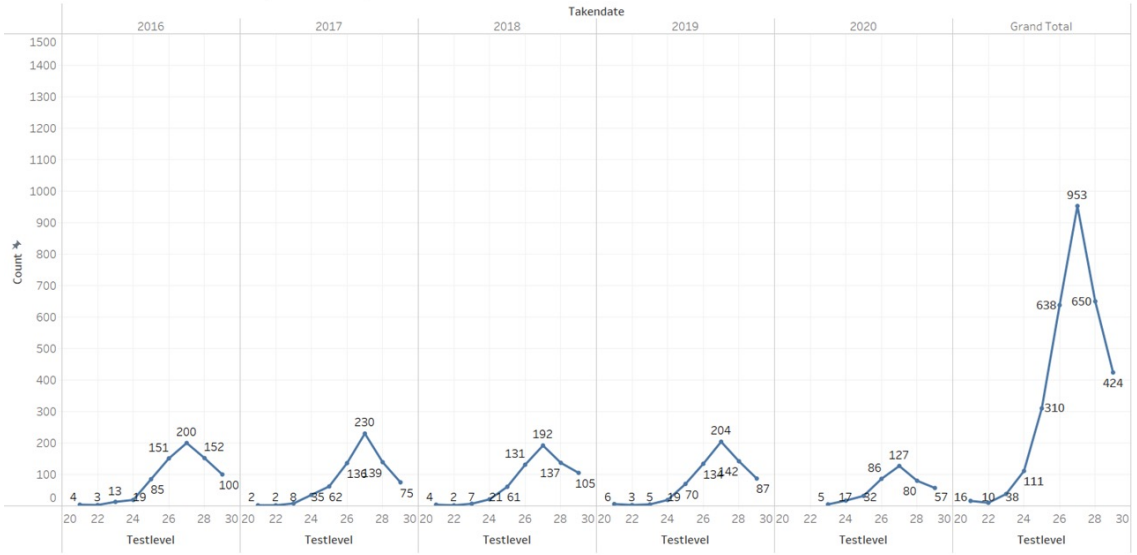
Count Vs TestLevel

Count Vs TakenDate **Grade 7 (1988-1992)**



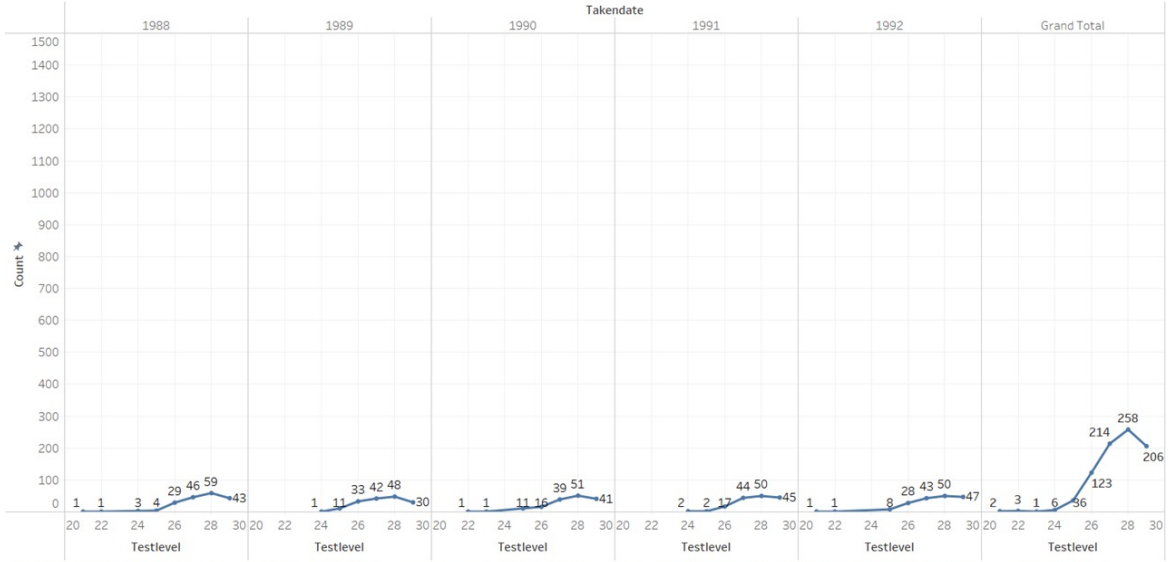
The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 7 to 7. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 1988, 1989, 1990, 1991 and 1992. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 395 members.

Count Vs TakenDate **Grade 7 (2016-2020)**



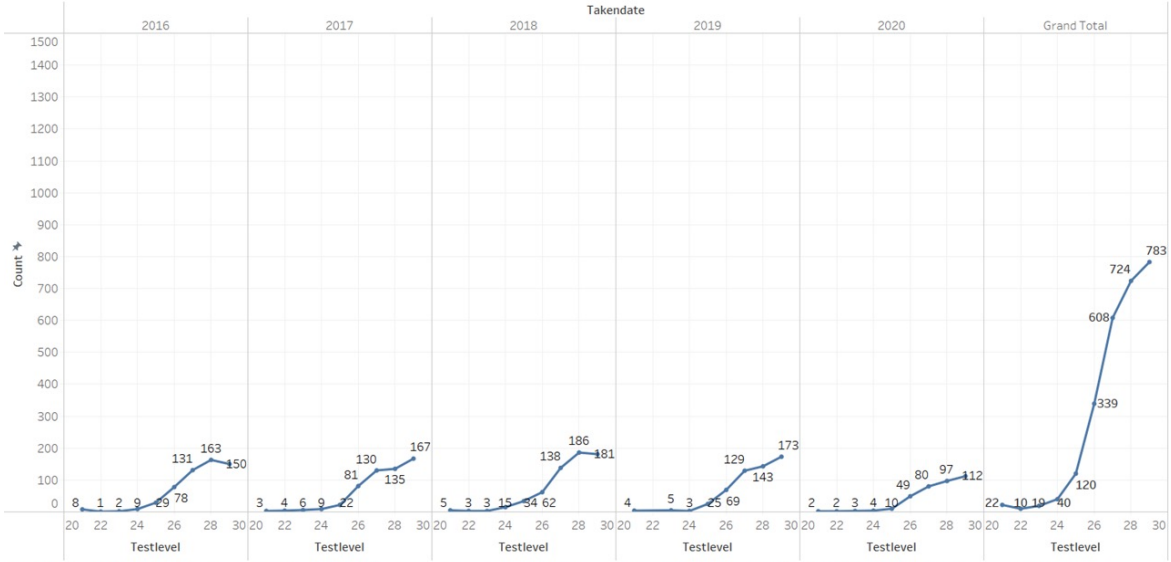
The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 7 to 7. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 2016, 2017, 2018, 2019 and 2020. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 396 members.

Count Vs TakenDate **Grade 8 (1988-1992)**



The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 8 to 8. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 1988, 1989, 1990, 1991 and 1992. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 395 members.

Count Vs TakenDate **Grade 8 (2015-2019)**

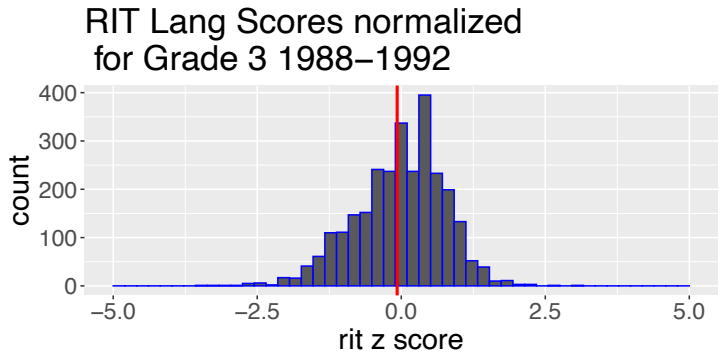


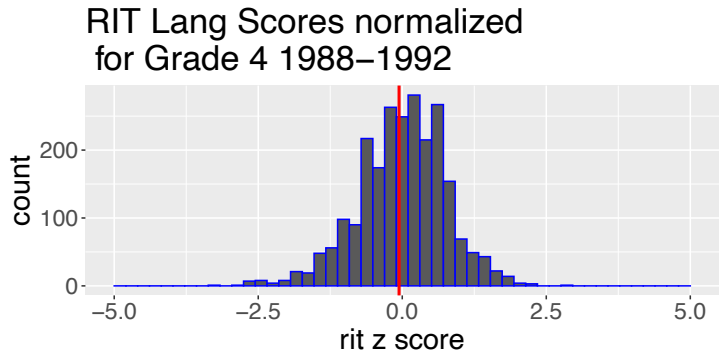
The trend of count of LanguageArts_All.csv for Testlevel broken down by Takendate Year. The data is filtered on Grade, which ranges from 8 to 8. The view is filtered on Takendate Year and Exclusions (Testlevel,YEAR(Takendate)). The Takendate Year filter keeps 2016, 2017, 2018, 2019 and 2020. The Exclusions (Testlevel,YEAR(Takendate)) filter keeps 396 members.

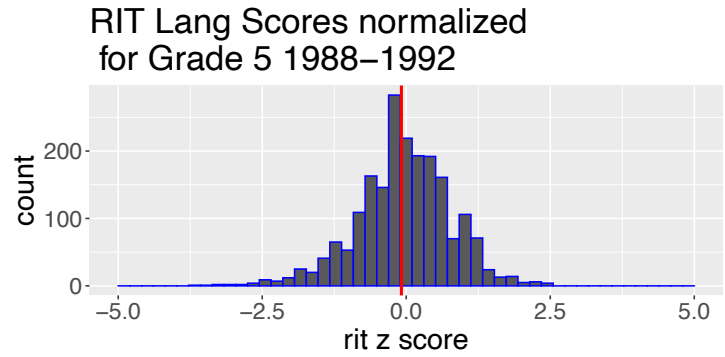
Distribution

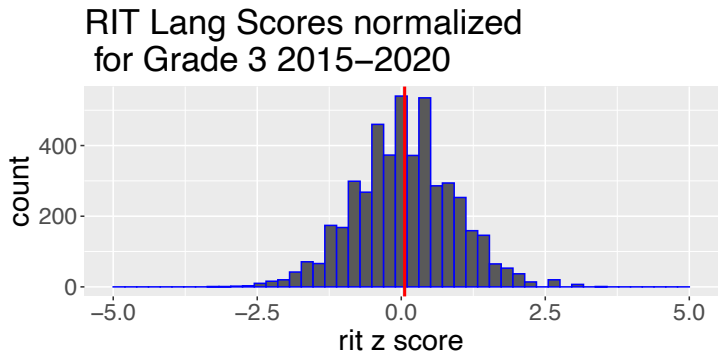
- mean RIT score has increased
- Lacking significant change in distribution
- Although students are doing better, the standard deviation is consistent

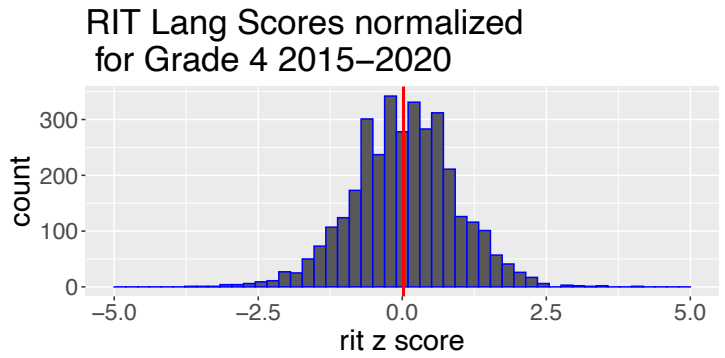
Distribution of Lang Scores (Early vs Late)

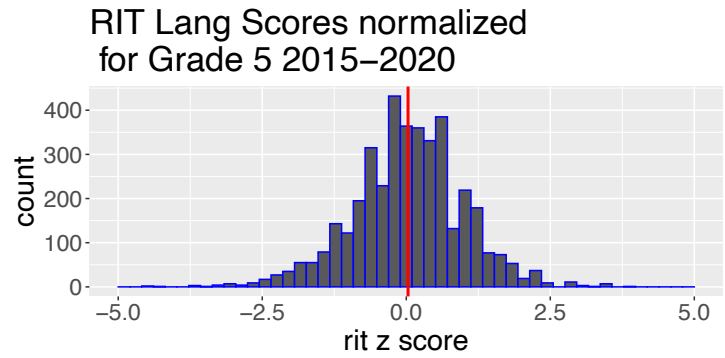






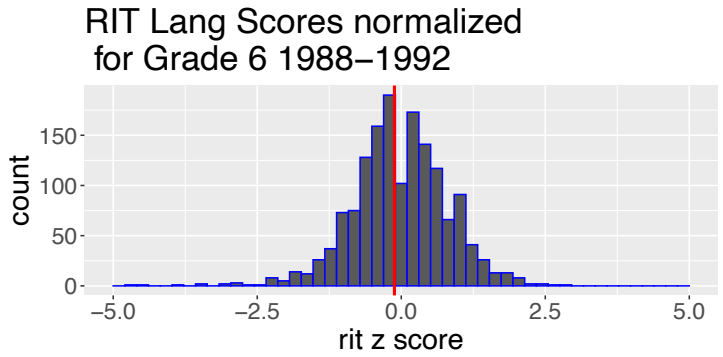


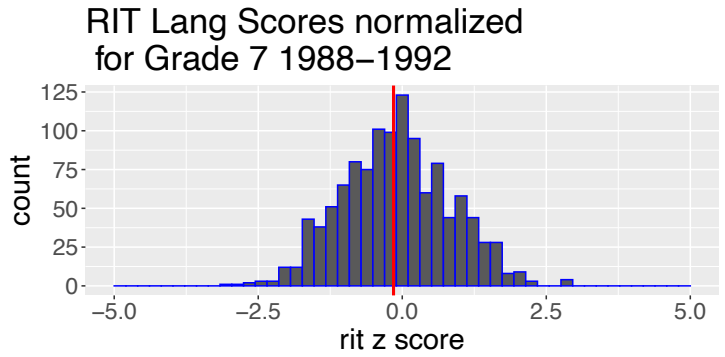


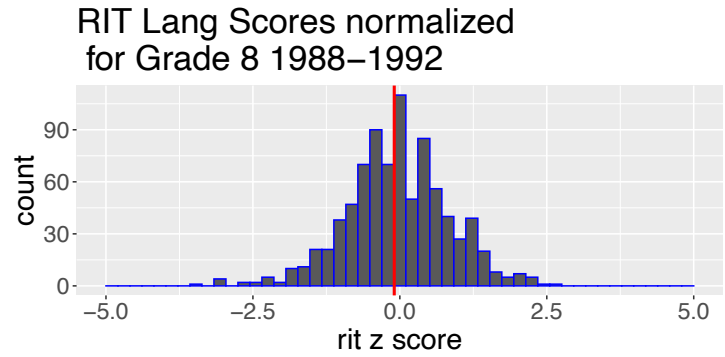


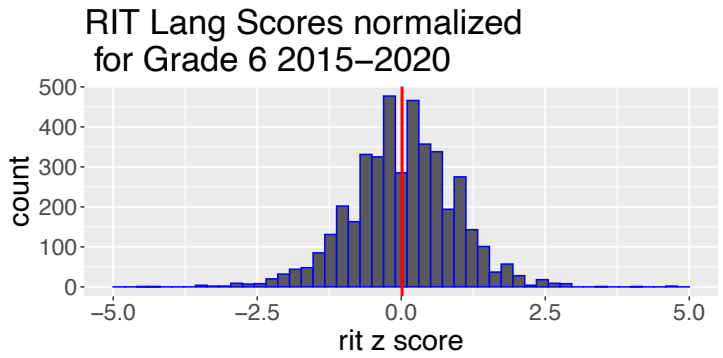
Conclusion: Distribution is relatively constant

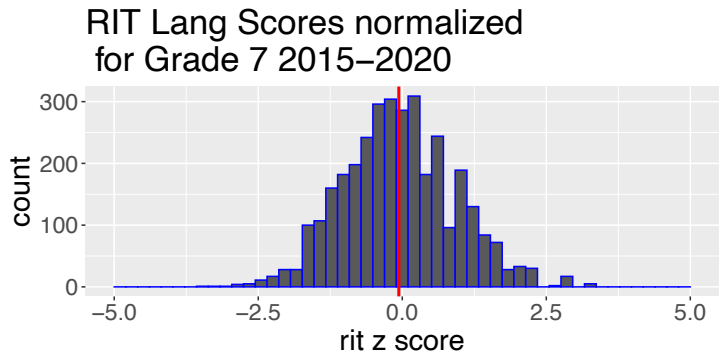
Distribution of Lang Scores (Early vs Late)

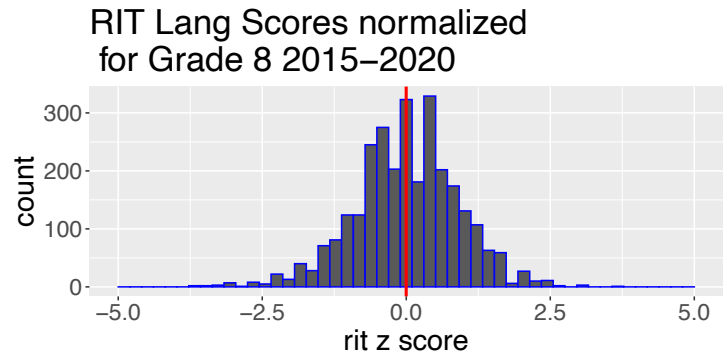






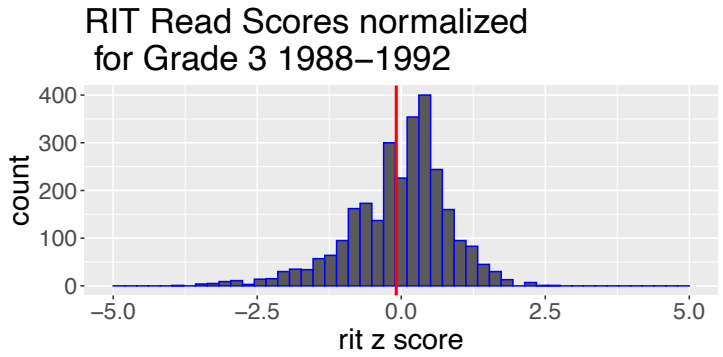


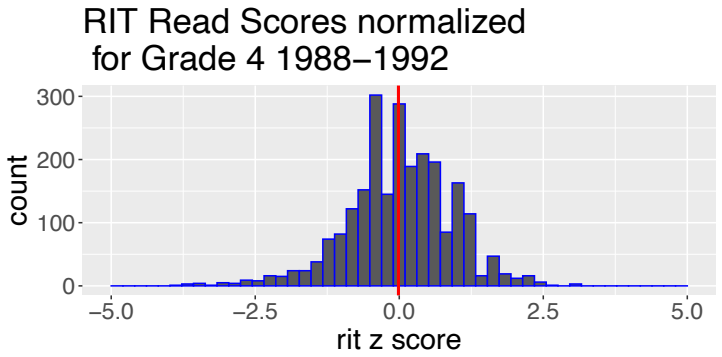


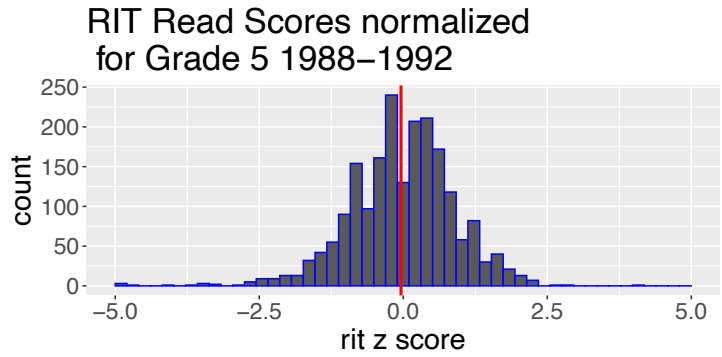


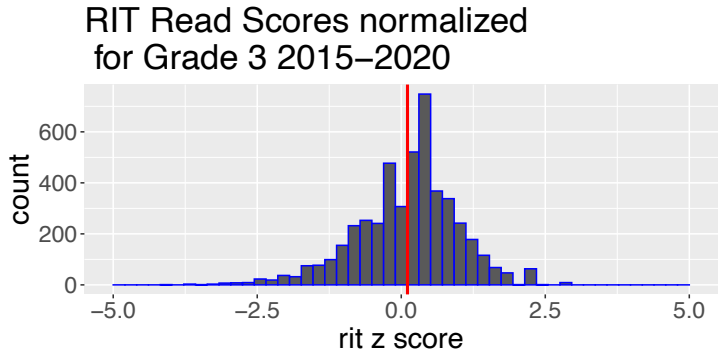
Conclusion: Distribution is relatively constant

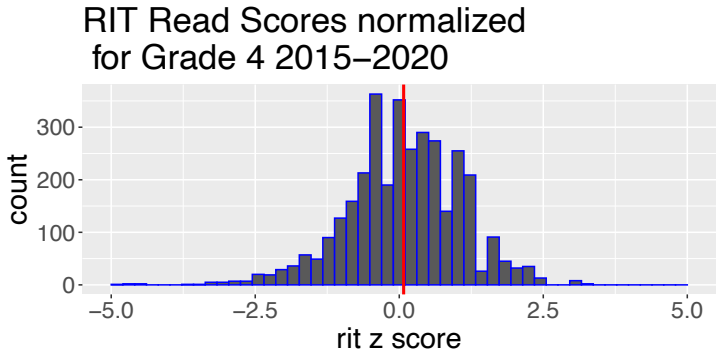
Distribution of Read Scores (Early vs Late)

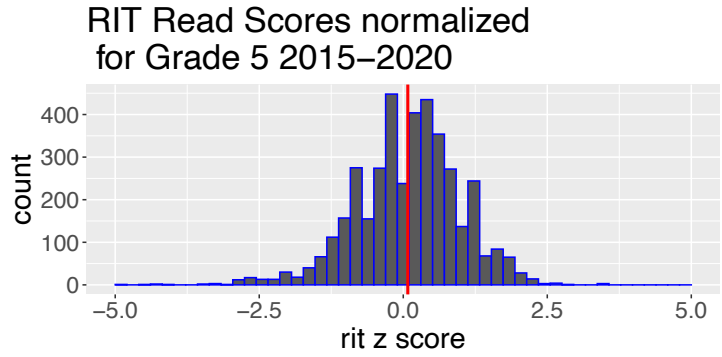






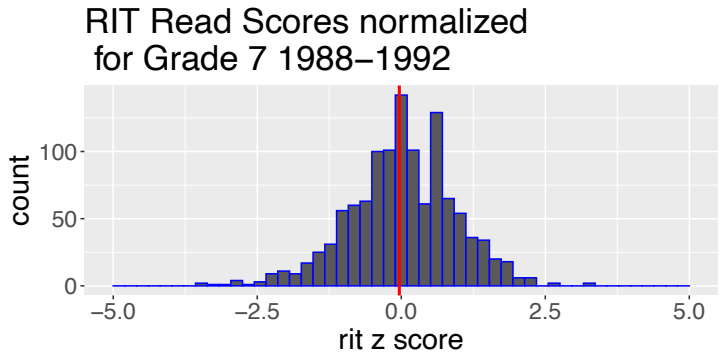


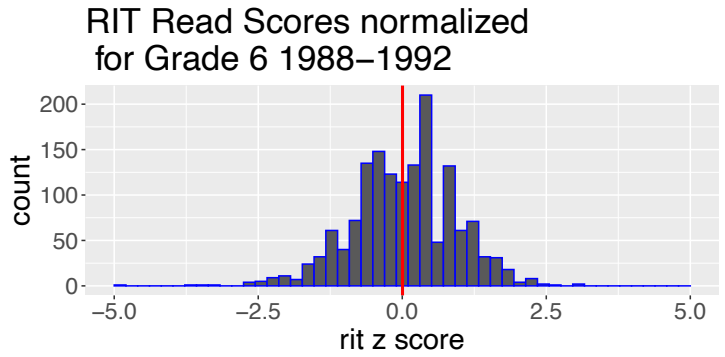


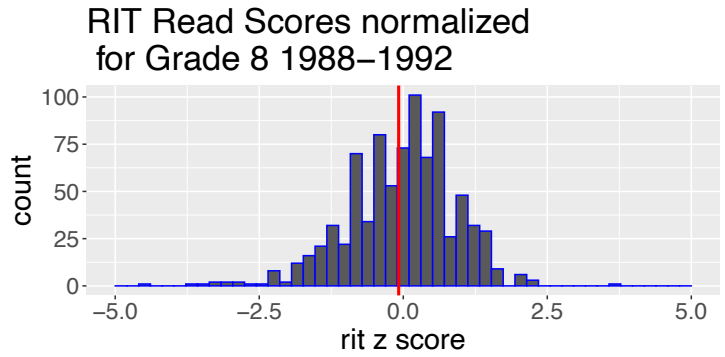


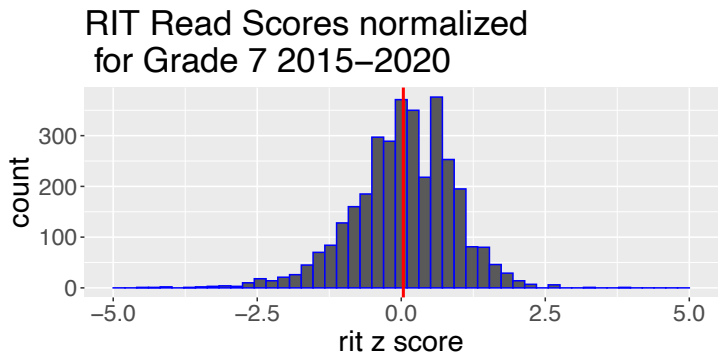
Conclusion: Distribution is relatively constant

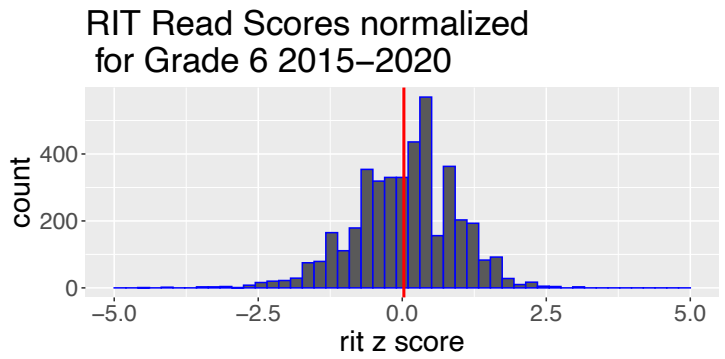
Distribution of Read Scores (Early vs Late)

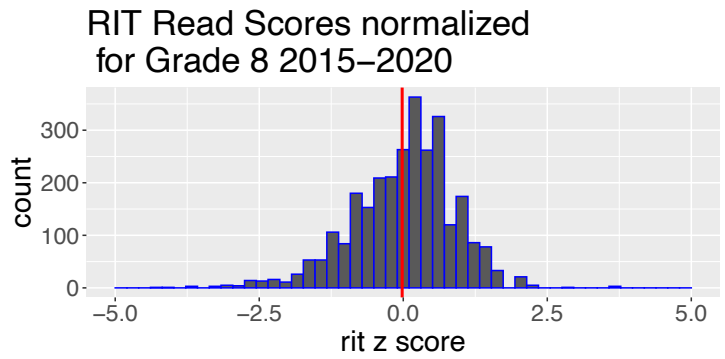






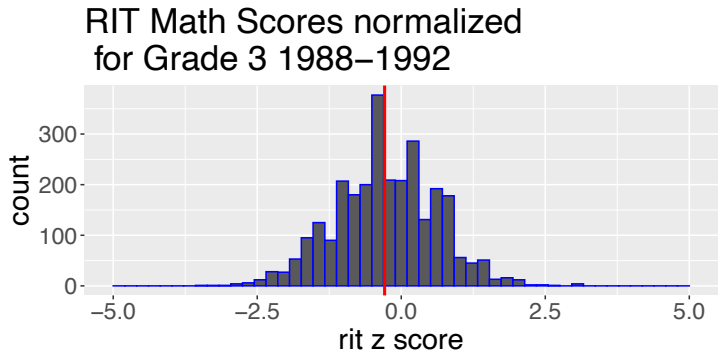


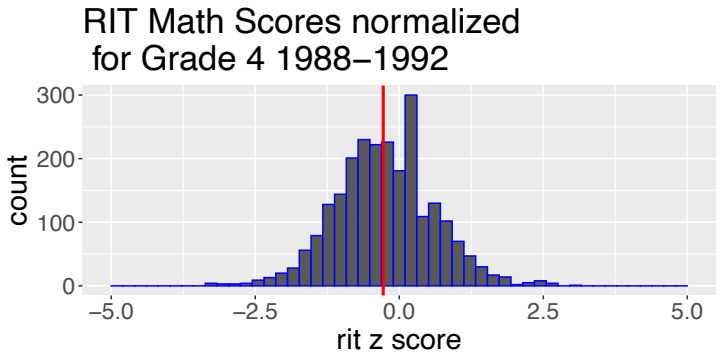


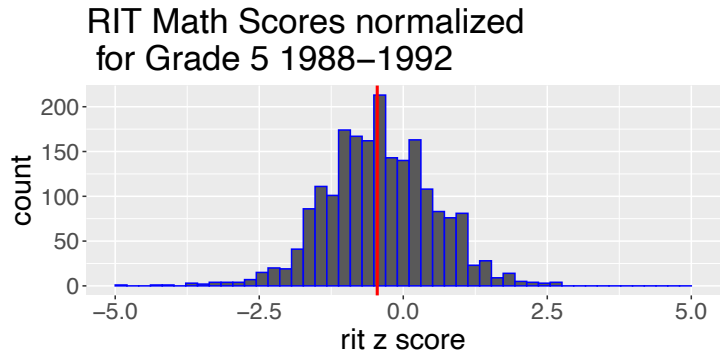


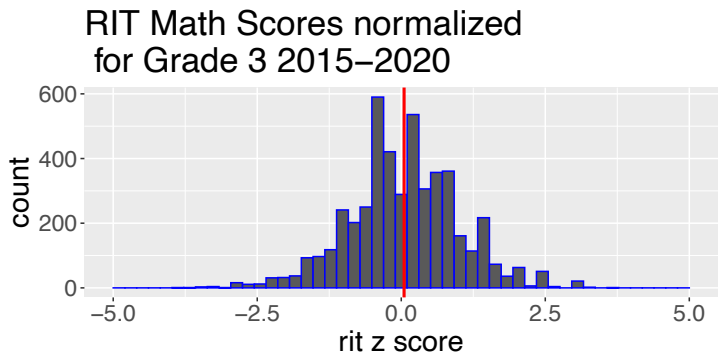
Conclusion: Distribution is relatively constant

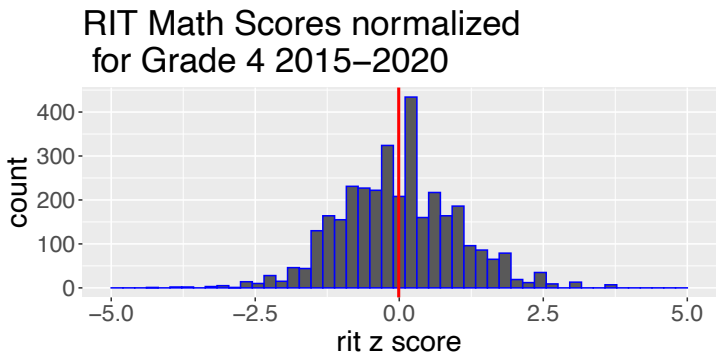
Distribution of Math Scores (Early vs Late)

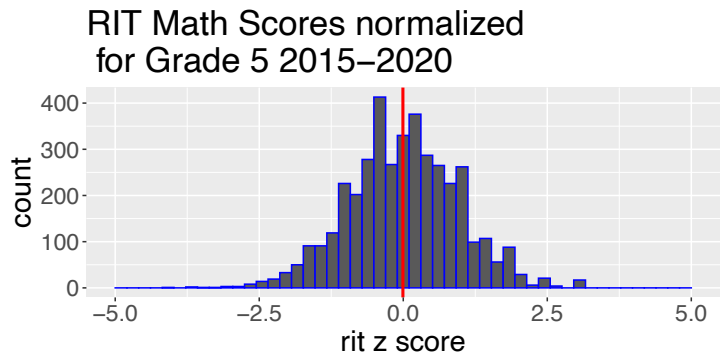






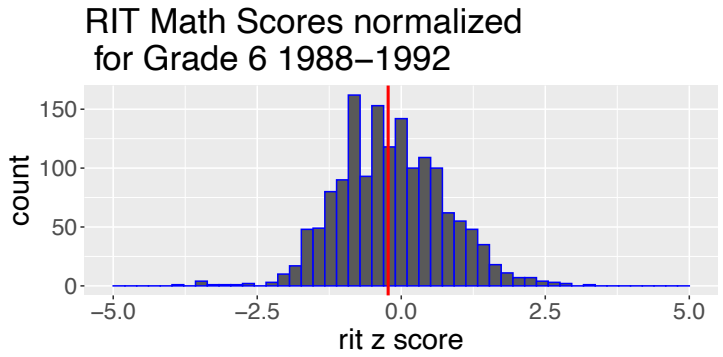


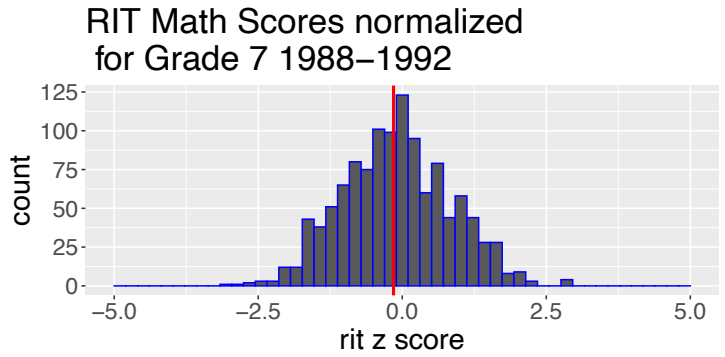


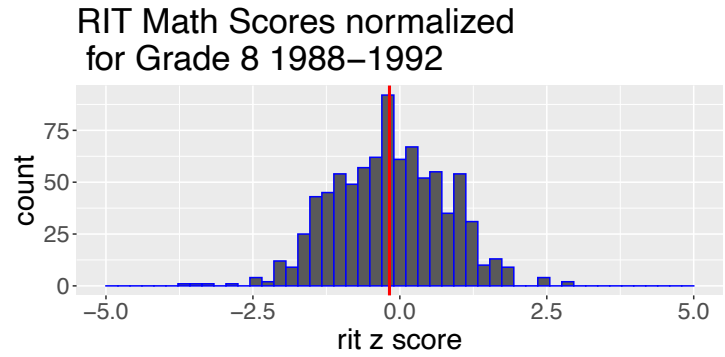


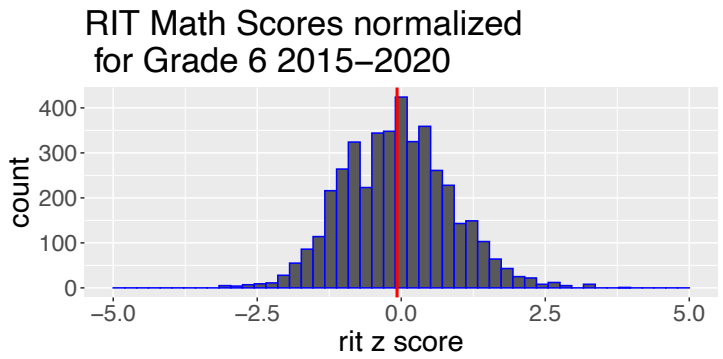
Conclusion: Distribution is relatively constant

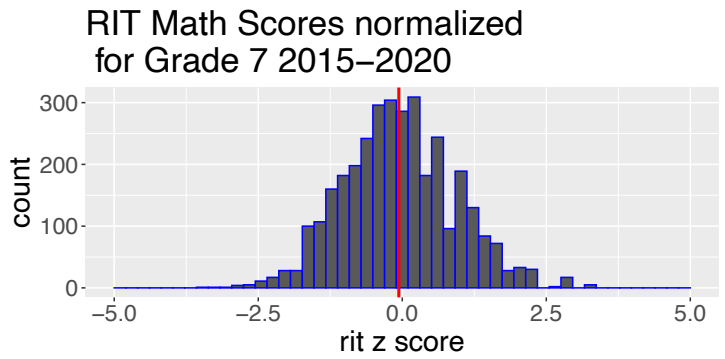
Distribution of Math Scores (Early vs Late)

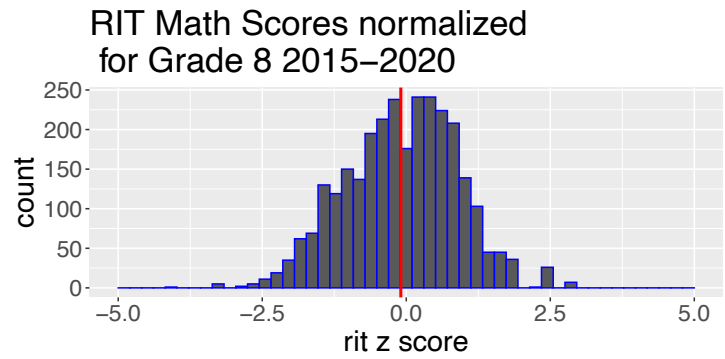






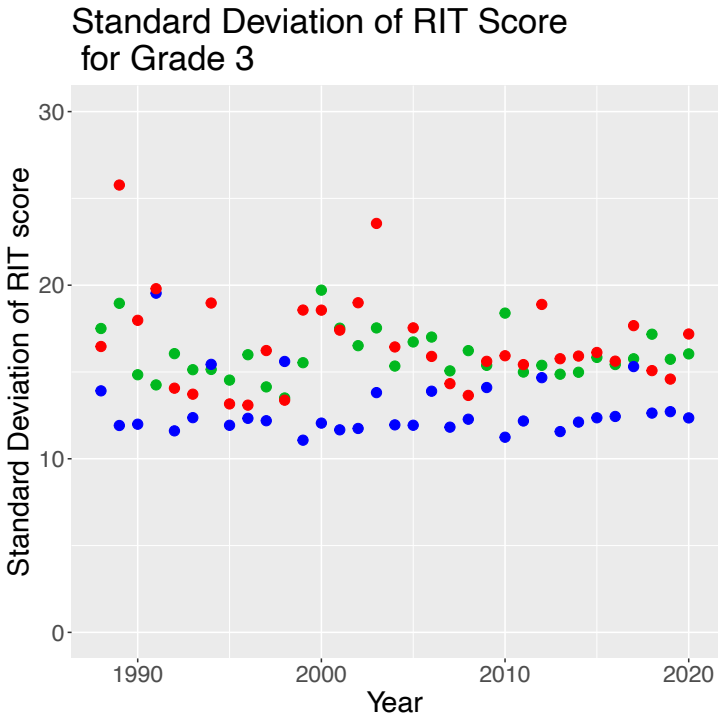


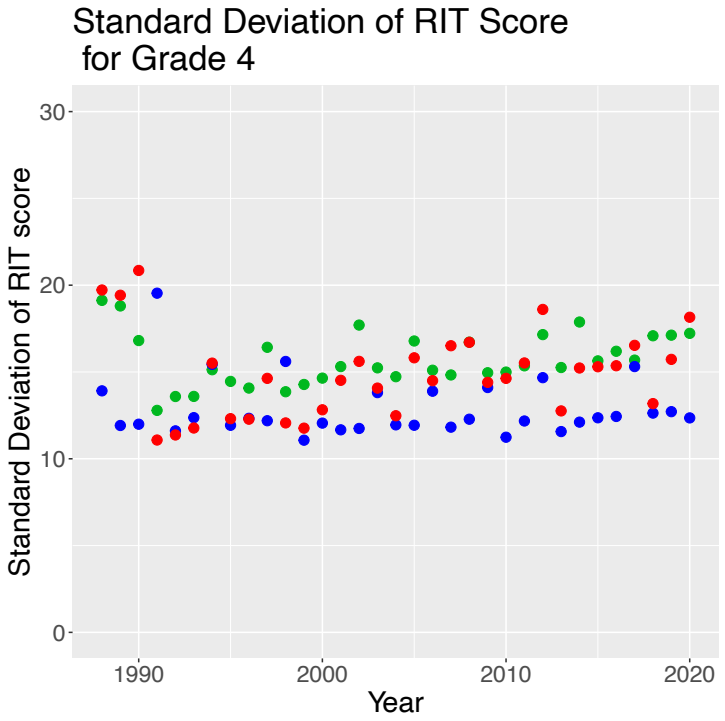


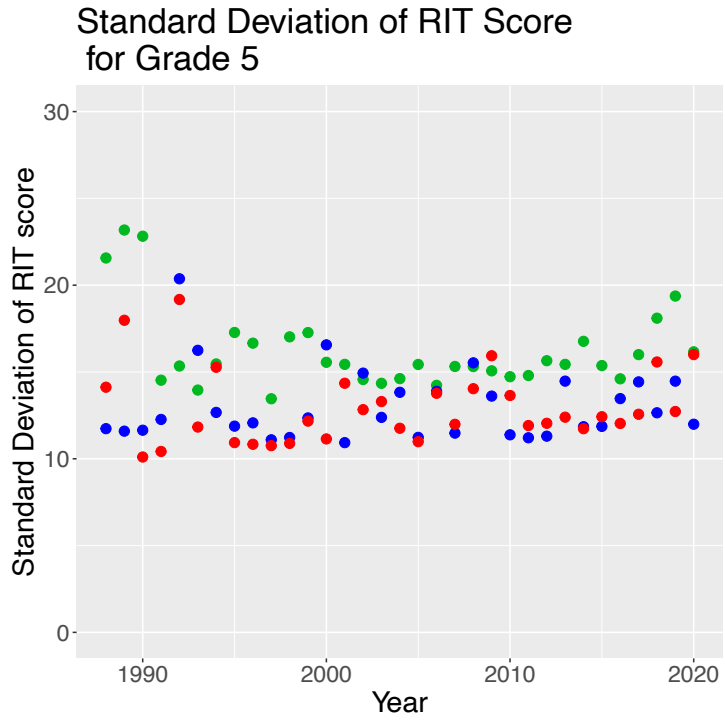


Conclusion: Distribution is relatively constant

Standard Deviation of RIT scores



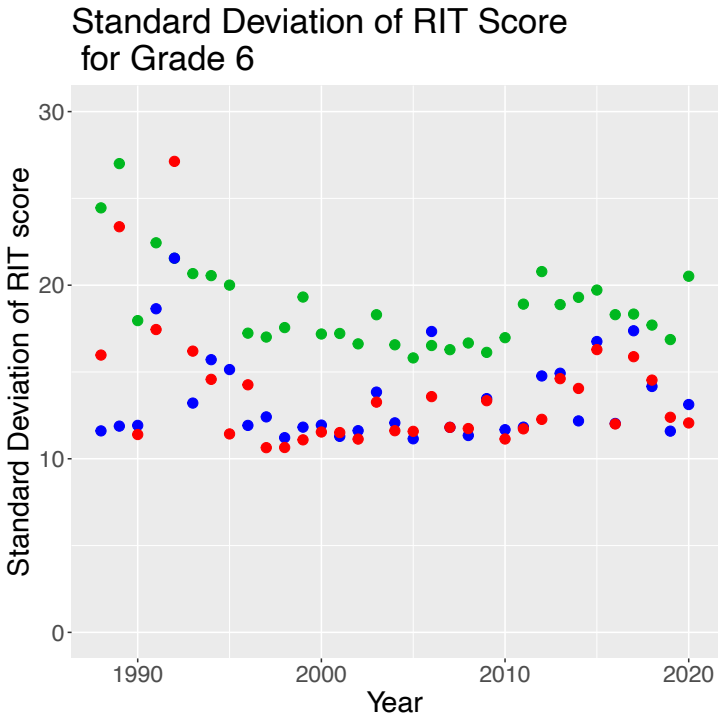


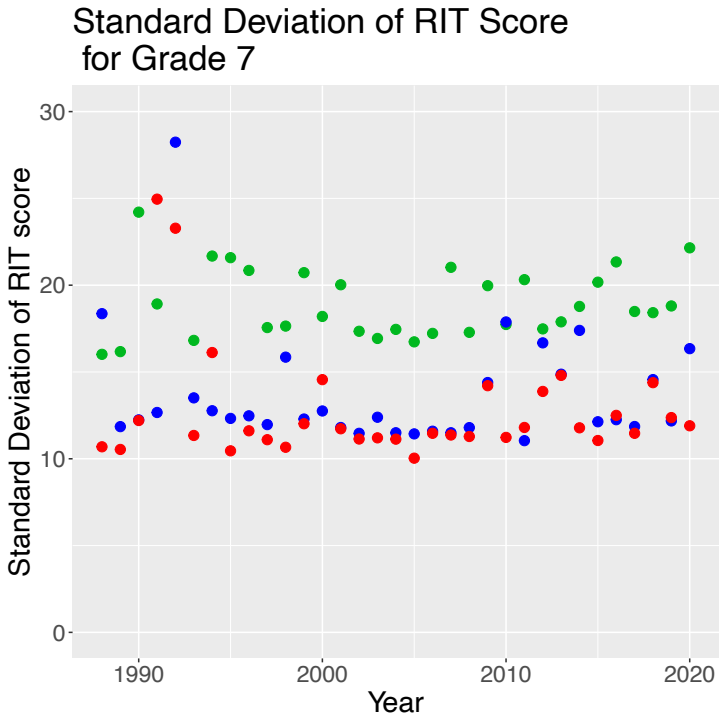


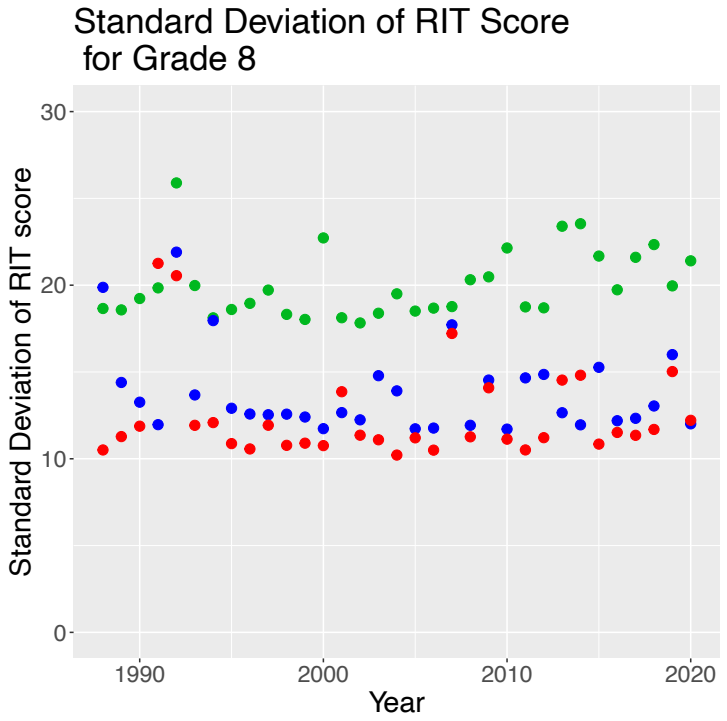
Reading
Math
Language

Conclusion: Distribution is relatively constant

Standard Deviation of RIT Scores







Reading
Math
Language

Conclusion: Distribution is relatively constant

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 1.072806
> mean(stdev_rit_df$pts_from_mean_read)
[1] 2.040893
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 1.945195
```

3rd

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 1.257912
> mean(stdev_rit_df$pts_from_mean_read)
[1] 1.964316
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 1.239319
```

4th

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 1.683956
> mean(stdev_rit_df$pts_from_mean_read)
[1] 1.676665
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 1.572627
```

5th

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 1.899546
> mean(stdev_rit_df$pts_from_mean_read)
[1] 2.41641
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 1.967825
```

6th

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 2.202976
> mean(stdev_rit_df$pts_from_mean_read)
[1] 2.079103
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 2.272348
```

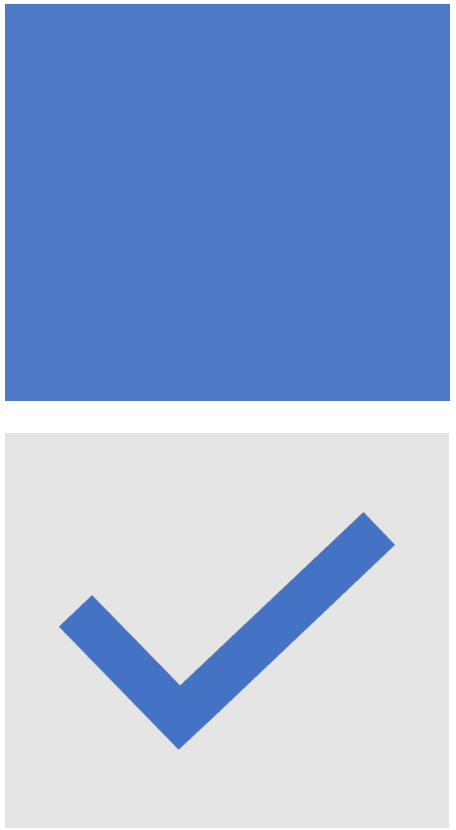
7th

```
> mean(stdev_rit_df$pts_from_mean_math)
[1] 1.536654
> mean(stdev_rit_df$pts_from_mean_read)
[1] 1.922609
> mean(stdev_rit_df$pts_from_mean_lang)
[1] 1.827273
```

8th

Error in Standard Deviation across Grades and Subjects

Modeling



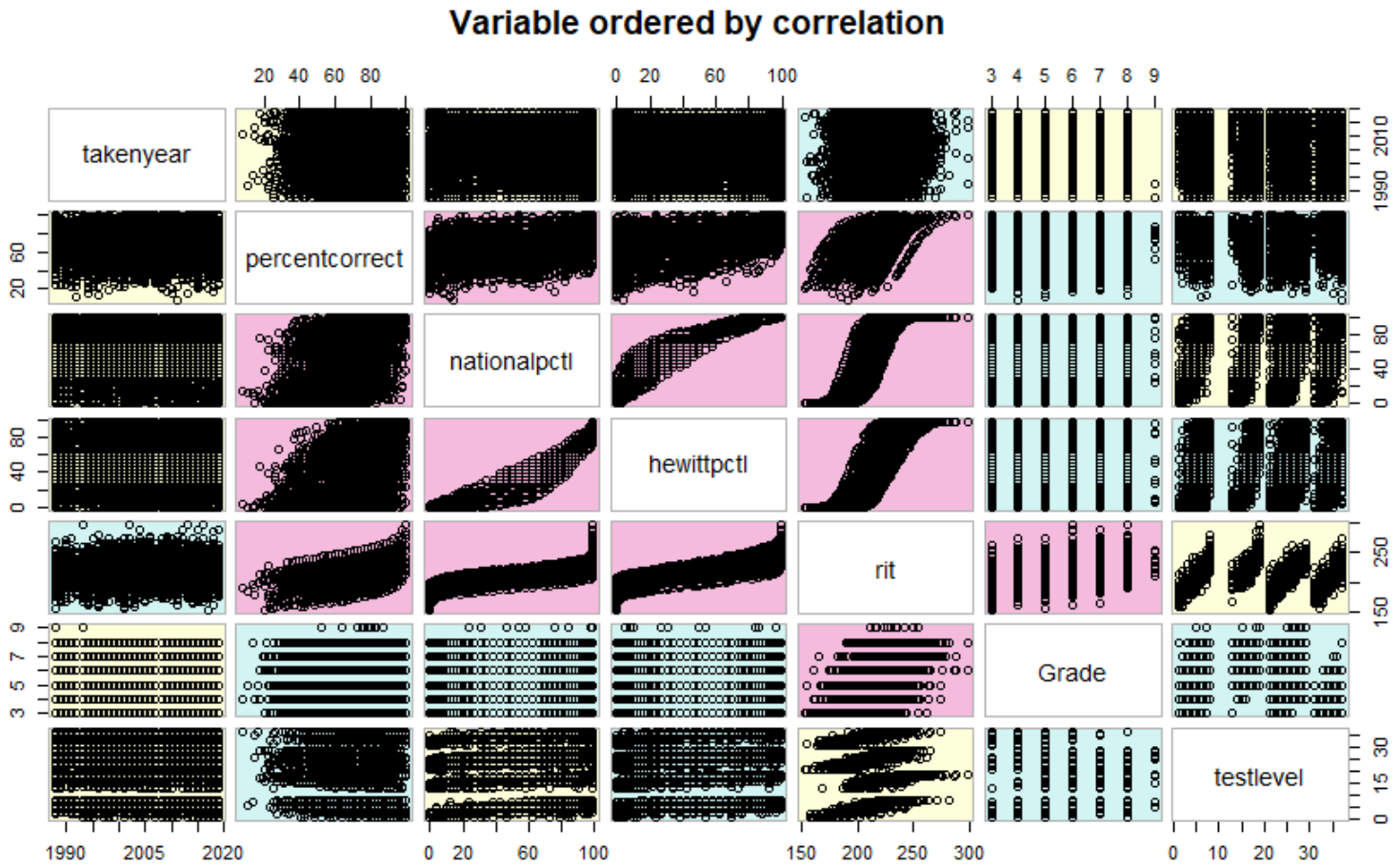
Choice of Linear Model

To identify the strength of our independent variable

Our data works well only with the linear regression modelling

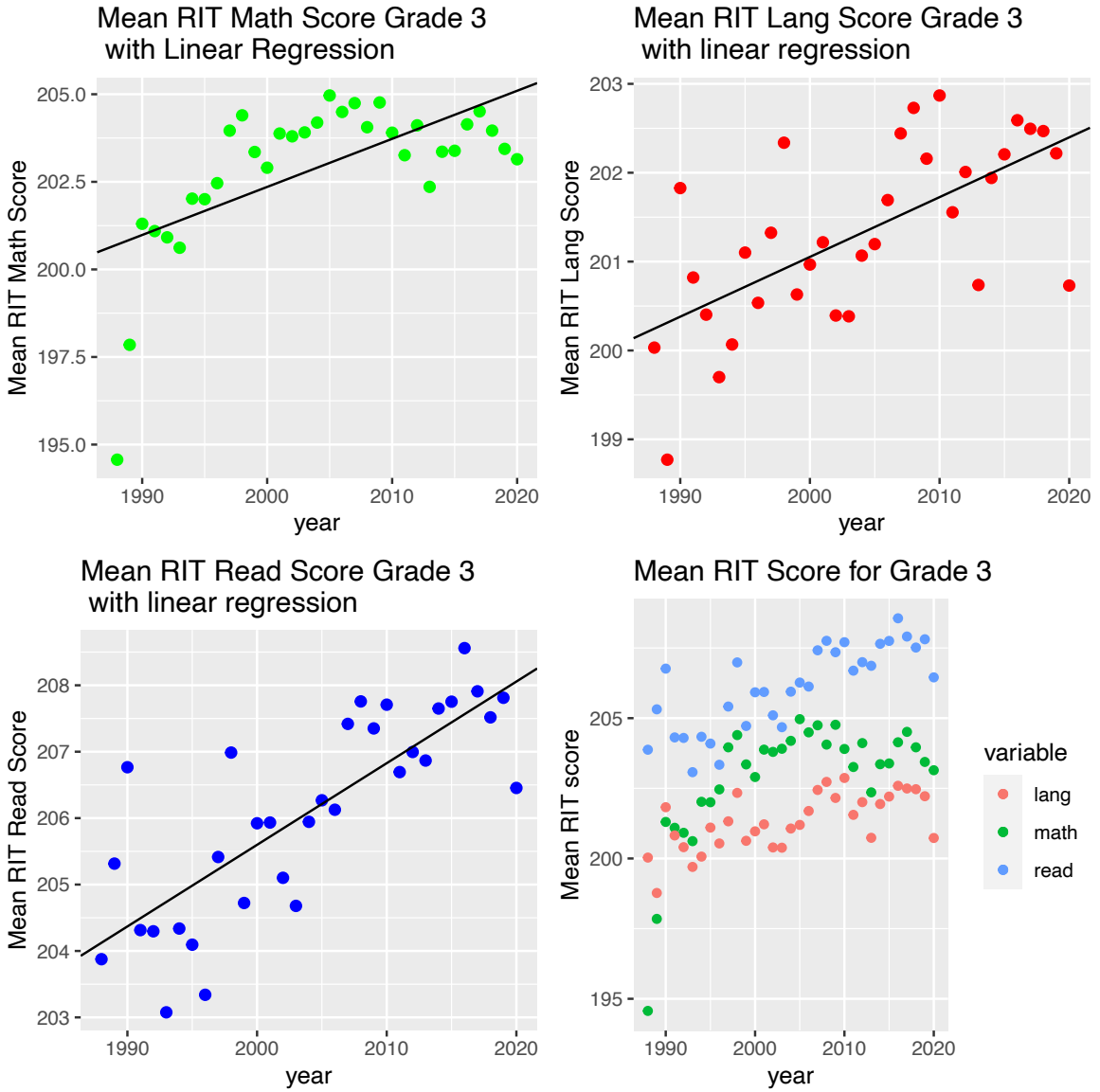
Other models are harder to understand with our data.

PCA Chart


Variable ordered by correlation
takenyear
percentcorrect
nationalpctl
hewittpctl
rit
Grade
testlevel

Modeling Mean RIT Score

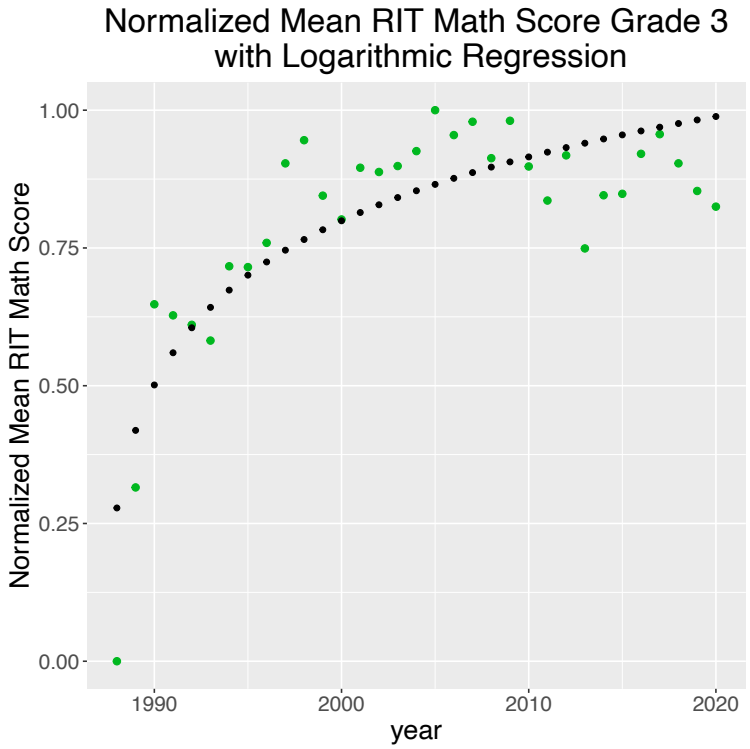
- Predicting mean RIT score from year + grade
- RIT score generalizes to average performance
- Normalized
- 33 data points – no test/train
- Less to predict and more to show trends (purpose of research)

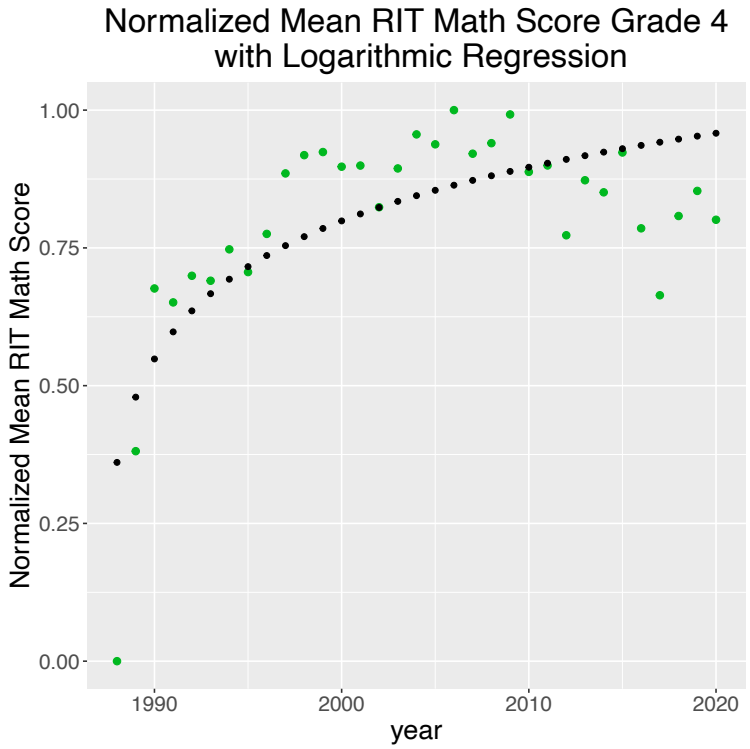


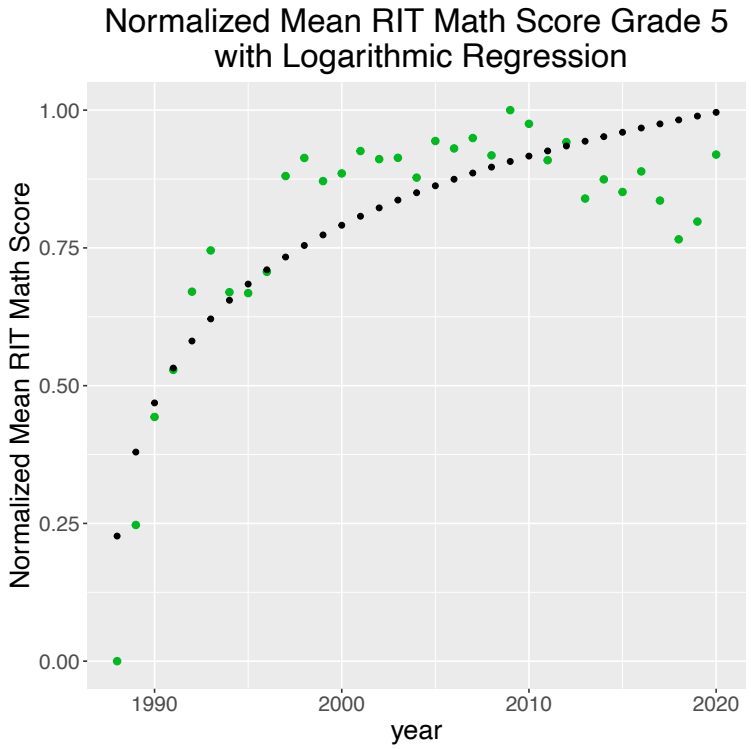
```r
# Predictors
math_mean_rit$norm_rit <-
  (math_mean_rit$mean_rit - min(math_mean_rit$mean_rit))/(max(math_mean_rit$mean_rit)-min(math_mean_rit$mean_rit))
math_mean_rit$norm_year <- math_mean_rit$year - (min(math_mean_rit$year) - 1)

# logarithmic regression math test
log_math_mean_model <- lm(formula = norm_rit ~ log(norm_year), data = math_mean_rit)
```

Mean Math RIT Score Prediction Model - Logarithmic

- better for the younger grades
- Demonstrates how math mean RIT score is plateauing

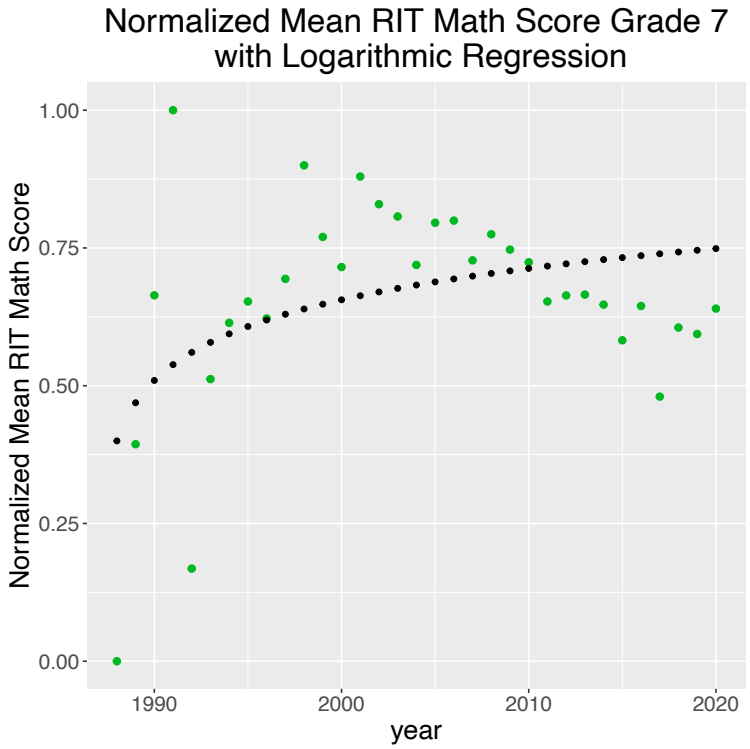


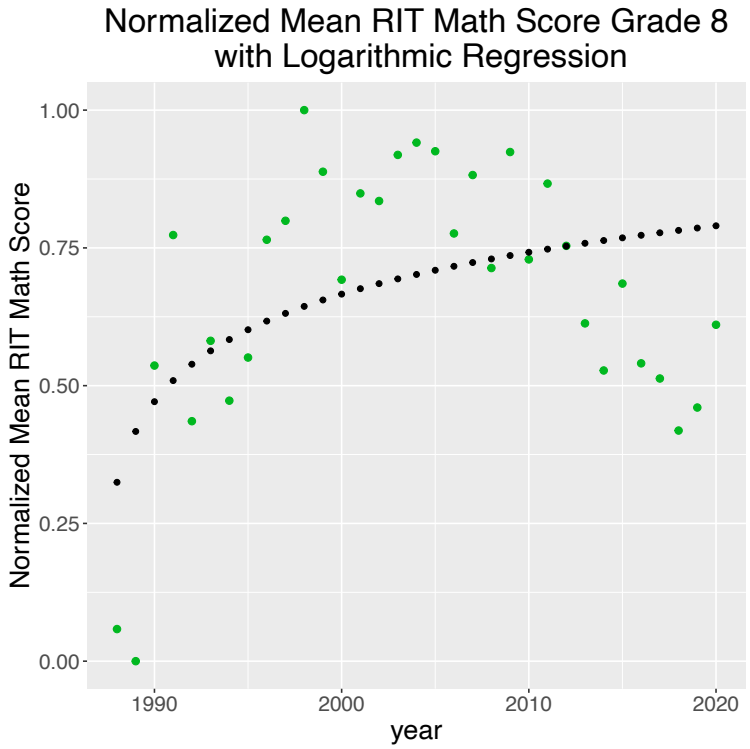


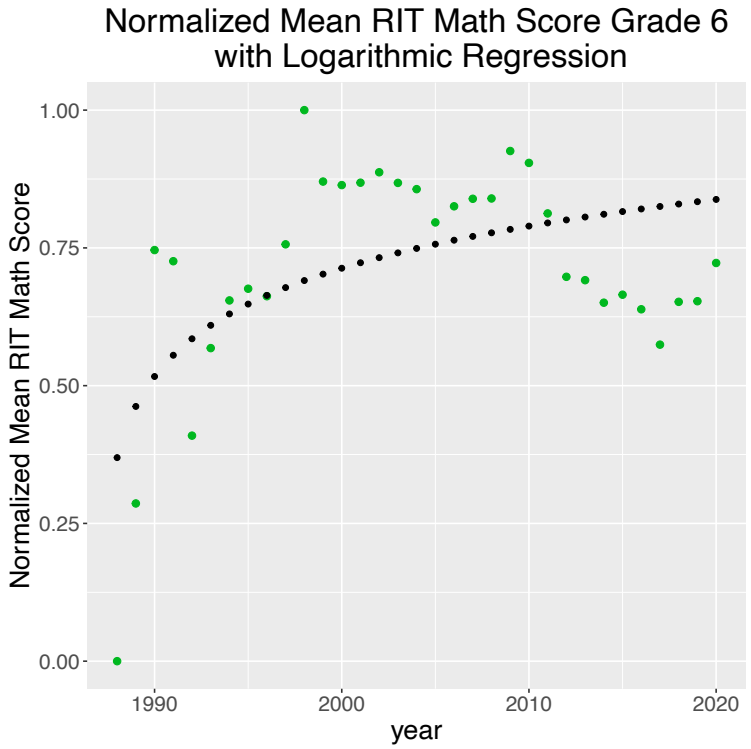


Adjusted R^2 Values for Each Grade

- 3rd - .73
- 4th - .58
- 5th - .75
- 6th - .32
- 7th - .19
- 8th - .23







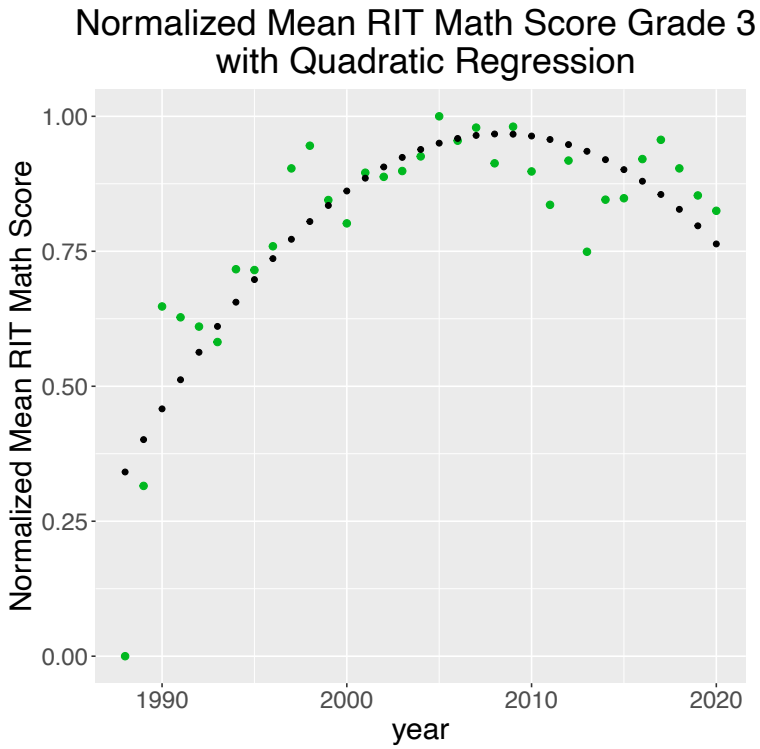
Adjusted R^2 Values for Each Grade

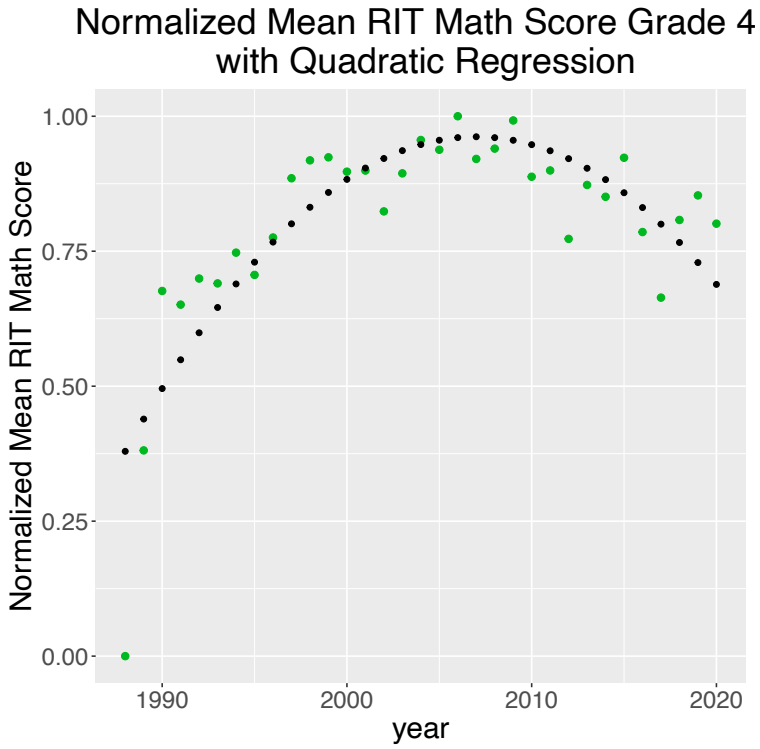
- 3rd - .73
- 4th - .58
- 5th - .75
- 6th - .32
- 7th - .19
- 8th - .23

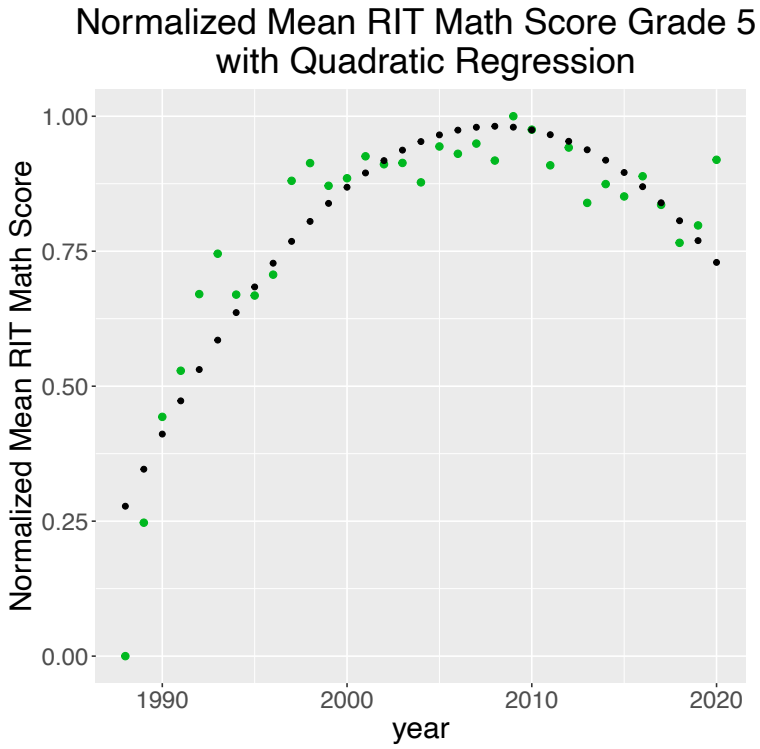
```r
# Predictors
math_mean_rit$norm_rit <-
  (math_mean_rit$mean_rit - min(math_mean_rit$mean_rit))/(max(math_mean_rit$mean_rit)-min(math_mean_rit$mean_rit))
math_mean_rit$norm_year <- math_mean_rit$year - (min(math_mean_rit$year) - 1)
math_mean_rit$norm_year_2 <- math_mean_rit$norm_year*math_mean_rit$norm_year

# logarithmic regression math test
quad_math_mean_model <- lm(formula = norm_rit ~ norm_year_2 + norm_year, data = math_mean_rit)
```

Mean Math RIT Score Prediction Model - Quadratic

- Better overall
- Average r^2 per grade: .66 vs .47
- Math scores are decreasing – returning back to original levels

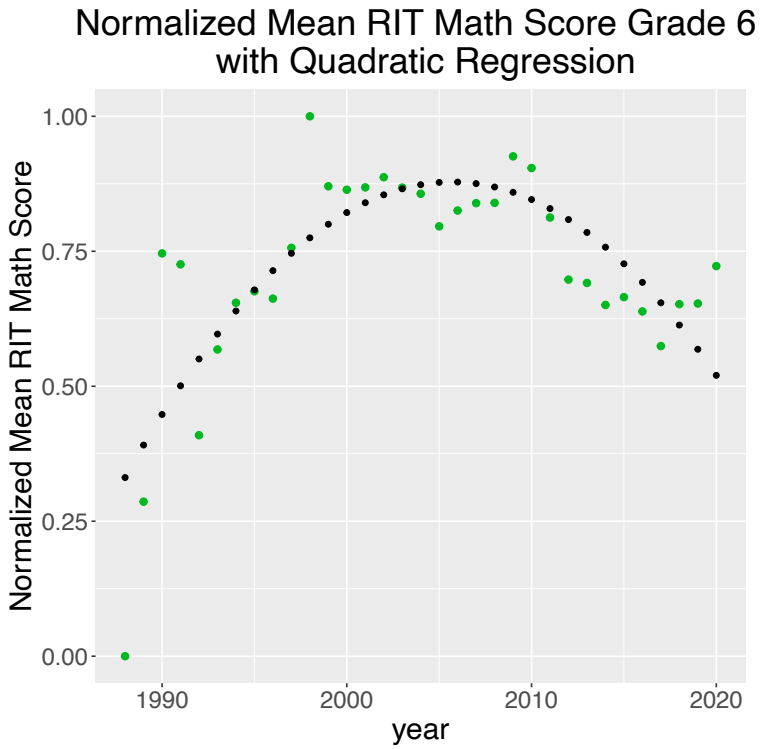


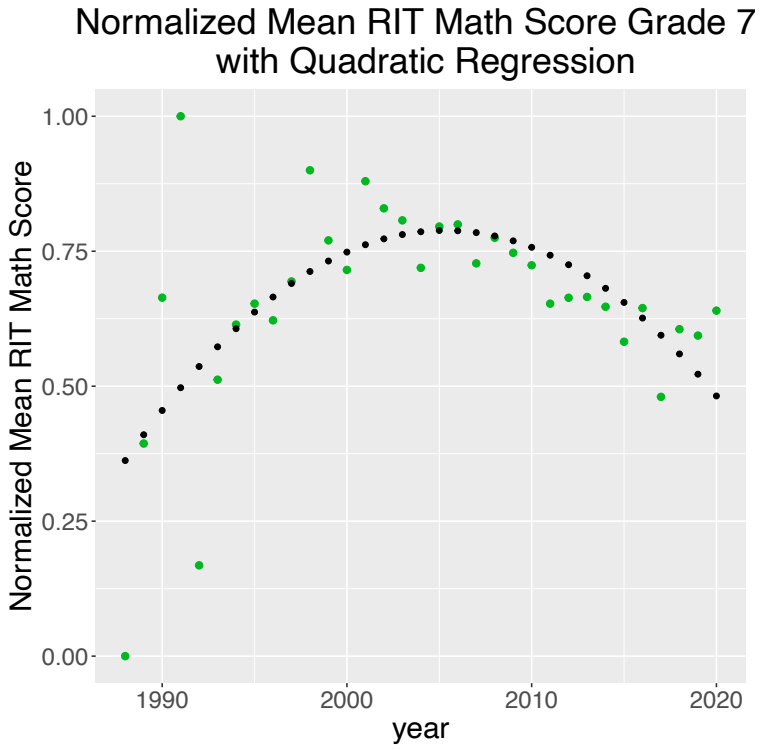


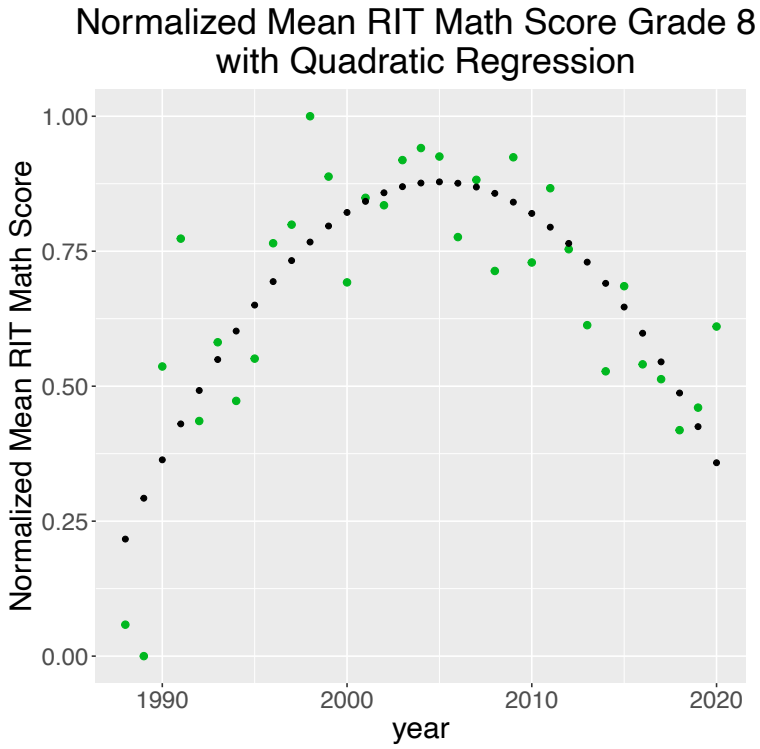


Adjusted R^2 Values for Each Grade

- 3rd - .75
- 4th - .72
- 5th - .84
- 6th - .60
- 7th - .36
- 8th - .68







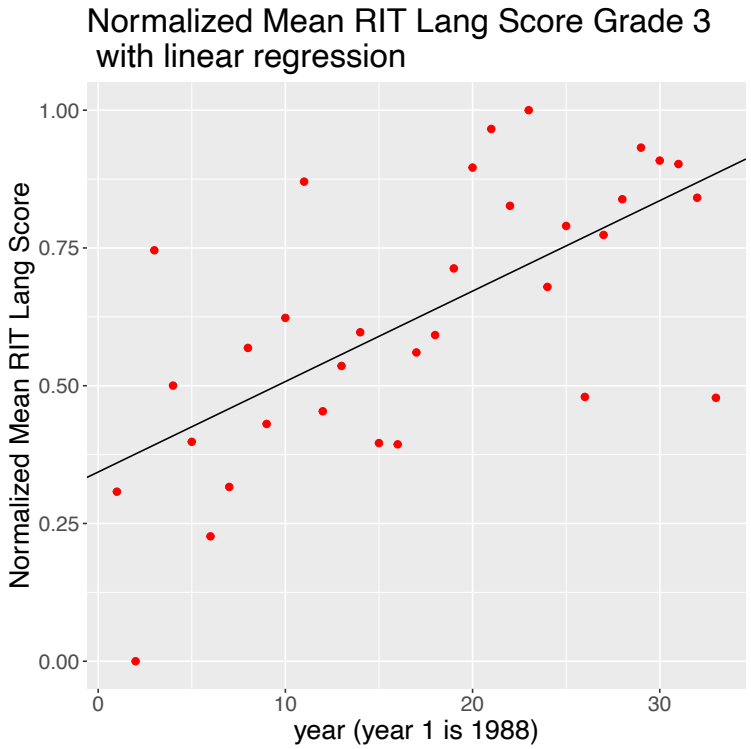
Adjusted R^2 Values for Each Grade

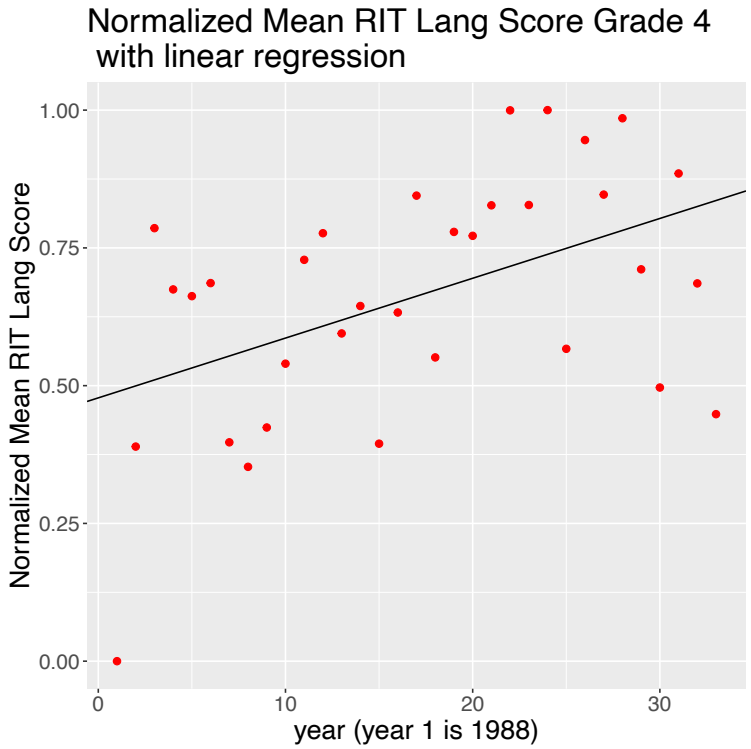
- 3rd - .75
- 4th - .72
- 5th - .84
- 6th - .60
- 7th - .36
- 8th - .68

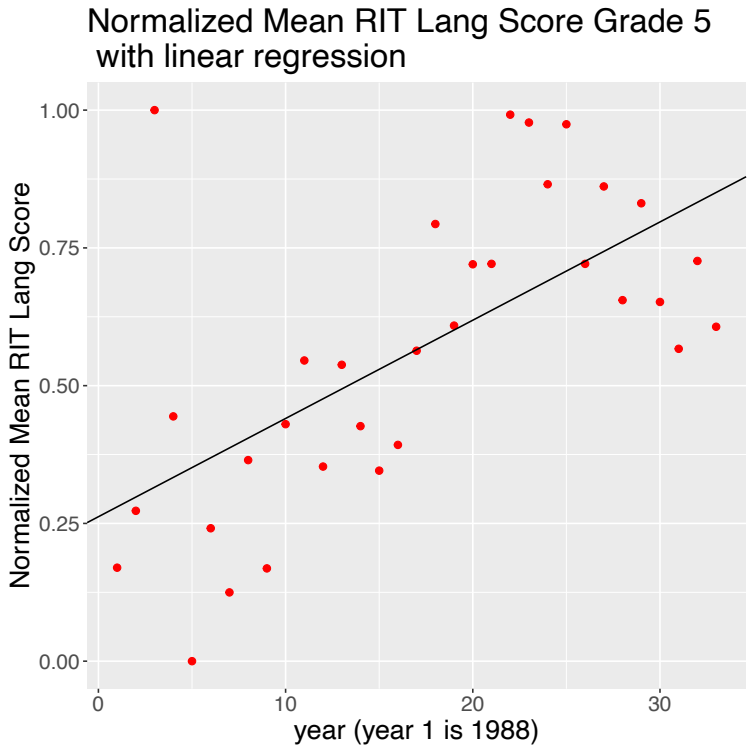
```
lang_mean_rit$norm_rit <-
  (lang_mean_rit$mean_rit - min(lang_mean_rit$mean_rit))/(max(lang_mean_rit$mean_rit)-min(lang_mean_rit$mean_rit))
lang_mean_rit$norm_year <- lang_mean_rit$year - (min(lang_mean_rit$year) - 1)

# liner regression model for math
# Double check our analysis using multi-variate model
lr_lang_mean = lm( formula = norm_rit ~ norm_year, data = lang_mean_rit)
summary(lr_lang_mean)
```

Mean RIT Language Prediction Model - Linear

- The model does not predict the Mean RIT scores well (quadratic and logarithmic were worse)
- Mean language scores have increased over the years
  - Students are doing better on the language test
  - Possible Reason: Their learning is more geared towards the material on the test?

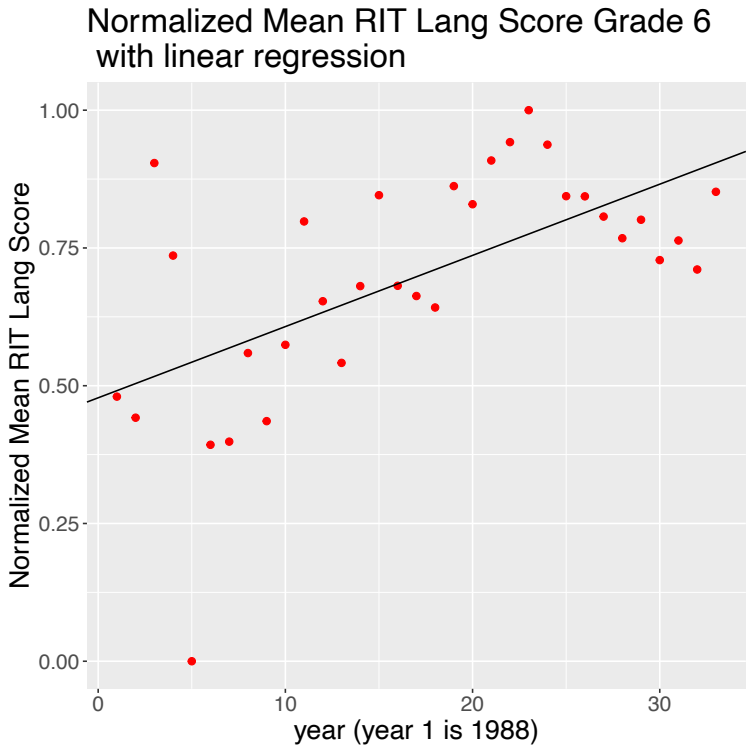


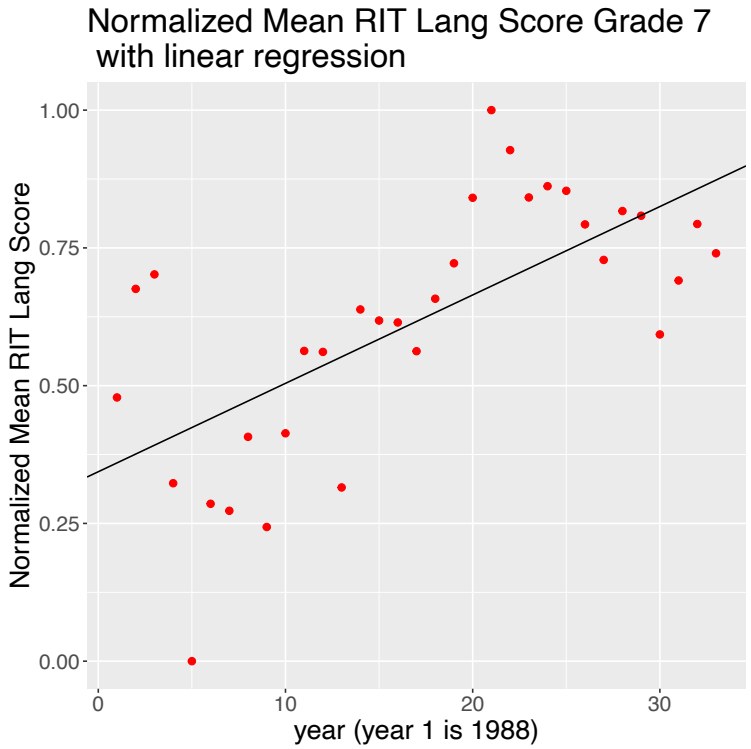


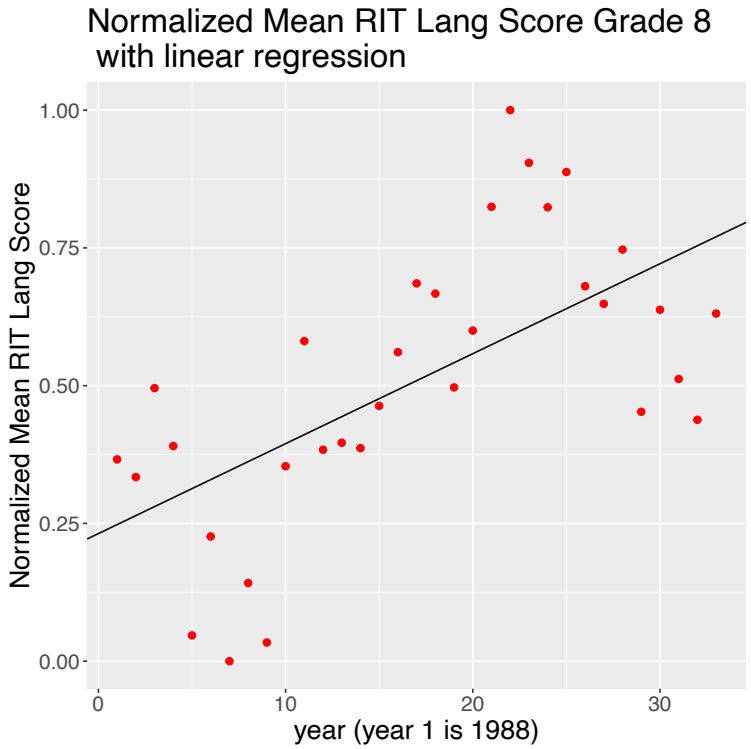


Mean Language RIT Score Prediction Model - Linear

- 3rd - .41
- 4th - .20
- 5th - .38
- 6th - .34
- 7th - .44
- 8th - .38







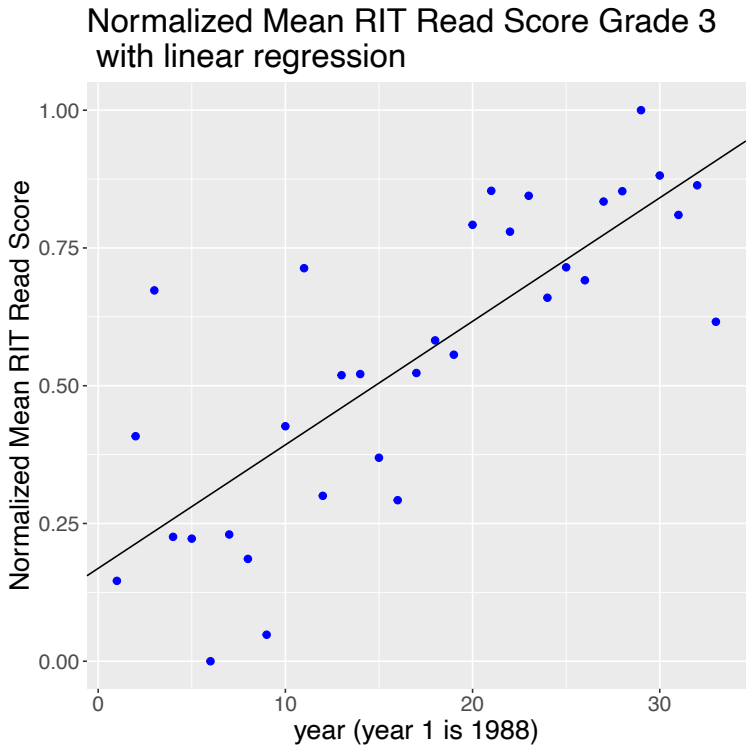
Mean Language RIT Score Prediction Model - Linear

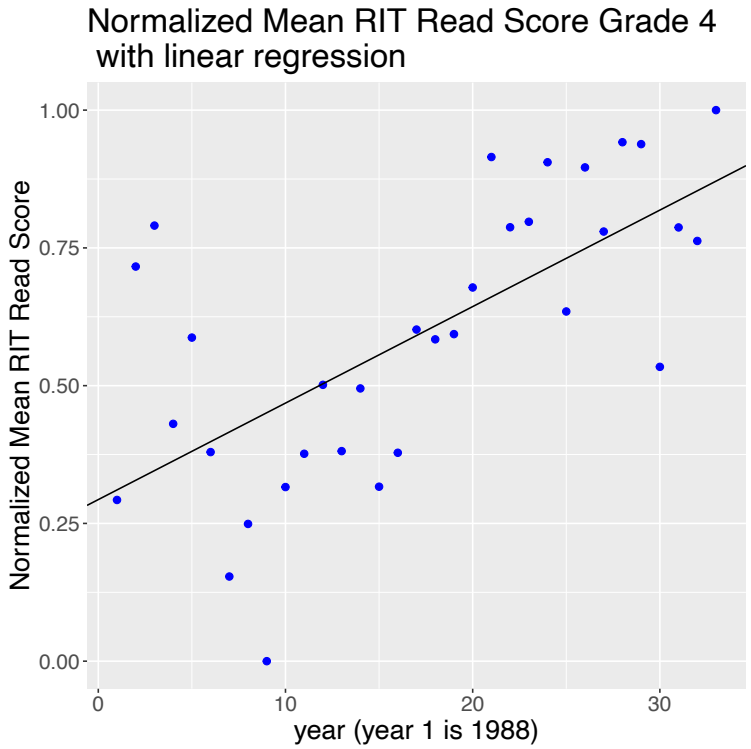
- 3rd - .41
- 4th - .20
- 5th - .38
- 6th - .34
- 7th - .44
- 8th - .38

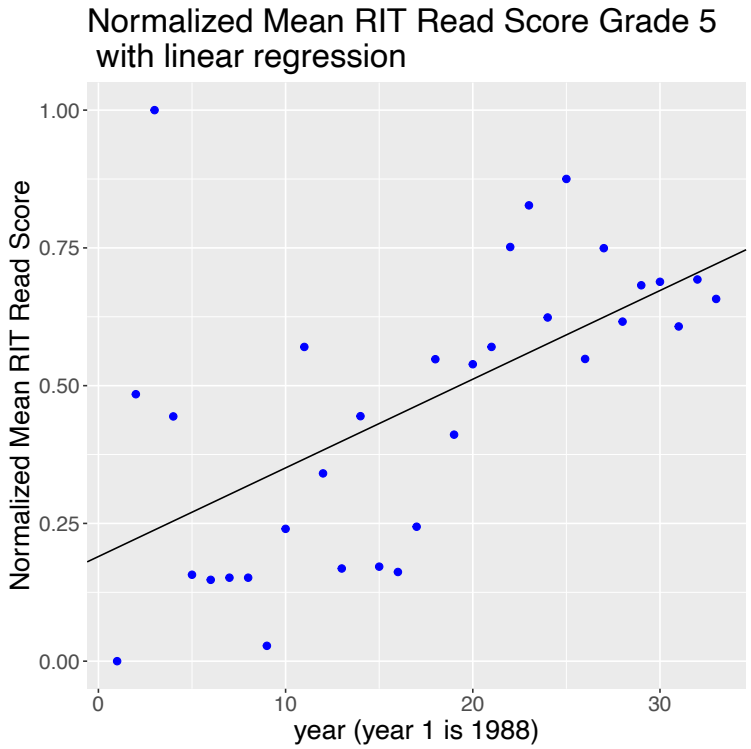
```r
# predictor and target variables
read_mean_rit$norm_rit <-
  (read_mean_rit$mean_rit - min(read_mean_rit$mean_rit))/(max(read_mean_rit$mean_r
read_mean_rit$norm_year <- read_mean_rit$year - (min(read_mean_rit$year) - 1)

# liner regression model for read
lr_read_mean = lm( formula = norm_rit ~ norm_year, data = read_mean_rit)
summary(lr_read_mean)
```

Mean RIT Reading Prediction Model - Linear

- Observation: Scores in early years are all over the place
  - Possible Reason: a standardized reading test was something very new to students, it has taken a while for them to get use to the test
- The model does not predict the Mean RIT scores well (quadratic and logarithmic were worse)
- Important: mean reading scores have increased over the years

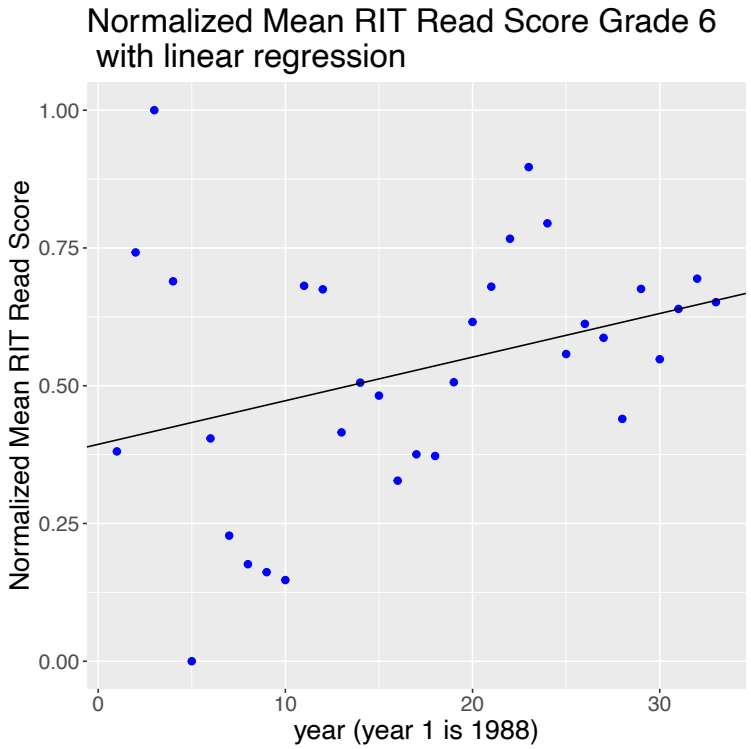


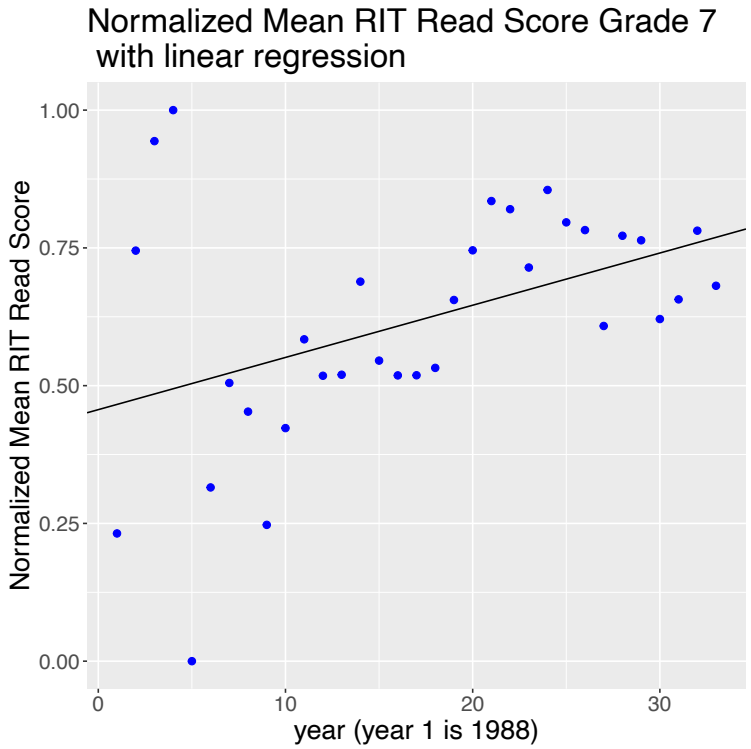


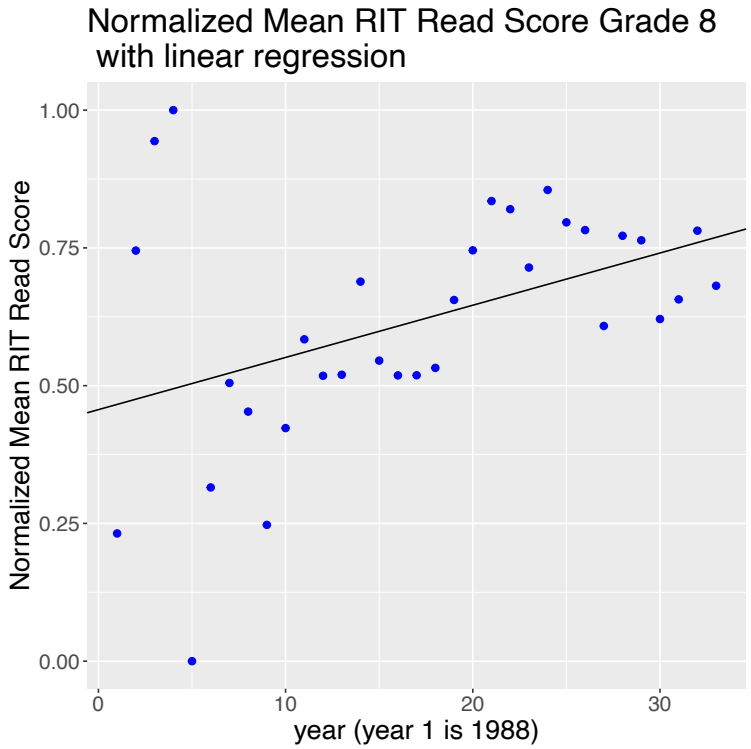


Mean Reading RIT Score Prediction Model - Linear

- 3rd - .61
- 4th - .42
- 5th - .34
- 6th - .10
- 7th - .15
- 8th - .13







Mean Reading RIT Score Prediction Model - Linear

- 3rd - .61
- 4th - .42
- 5th - .34
- 6th - .10
- 7th - .15
- 8th - .13

Linear Model for predicting mean RIT scores

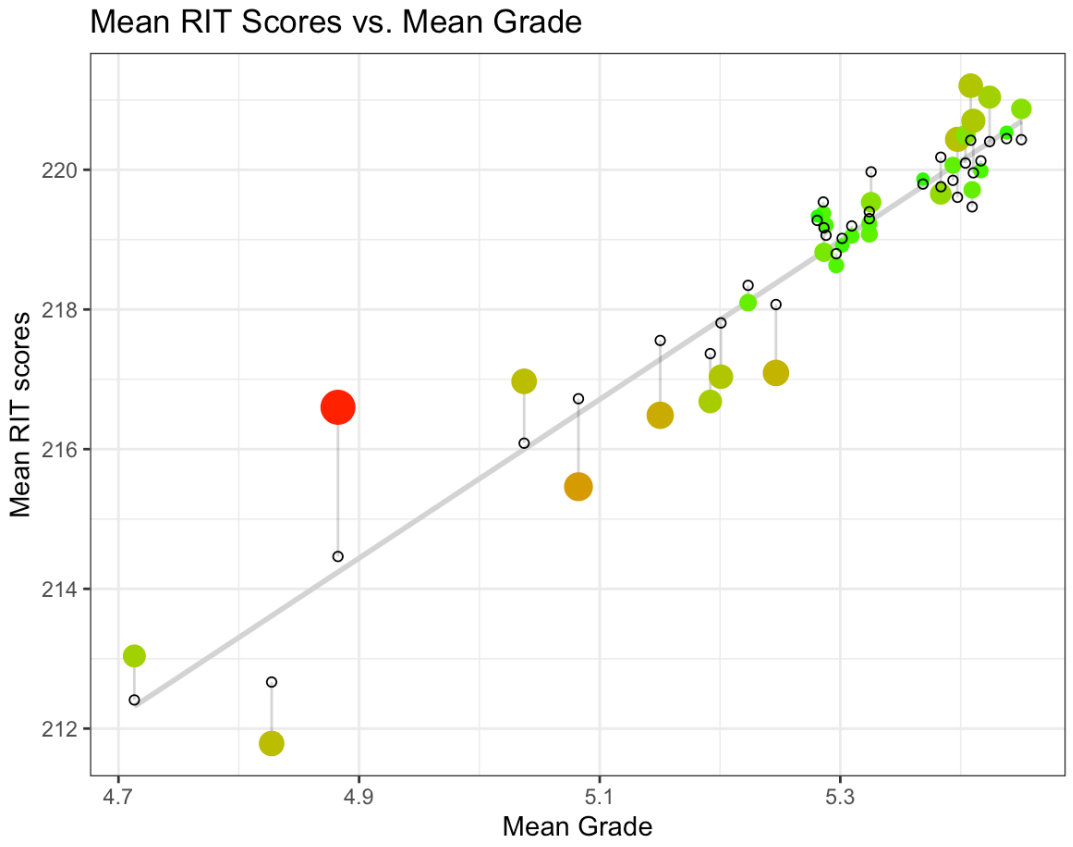
Grouping the Data

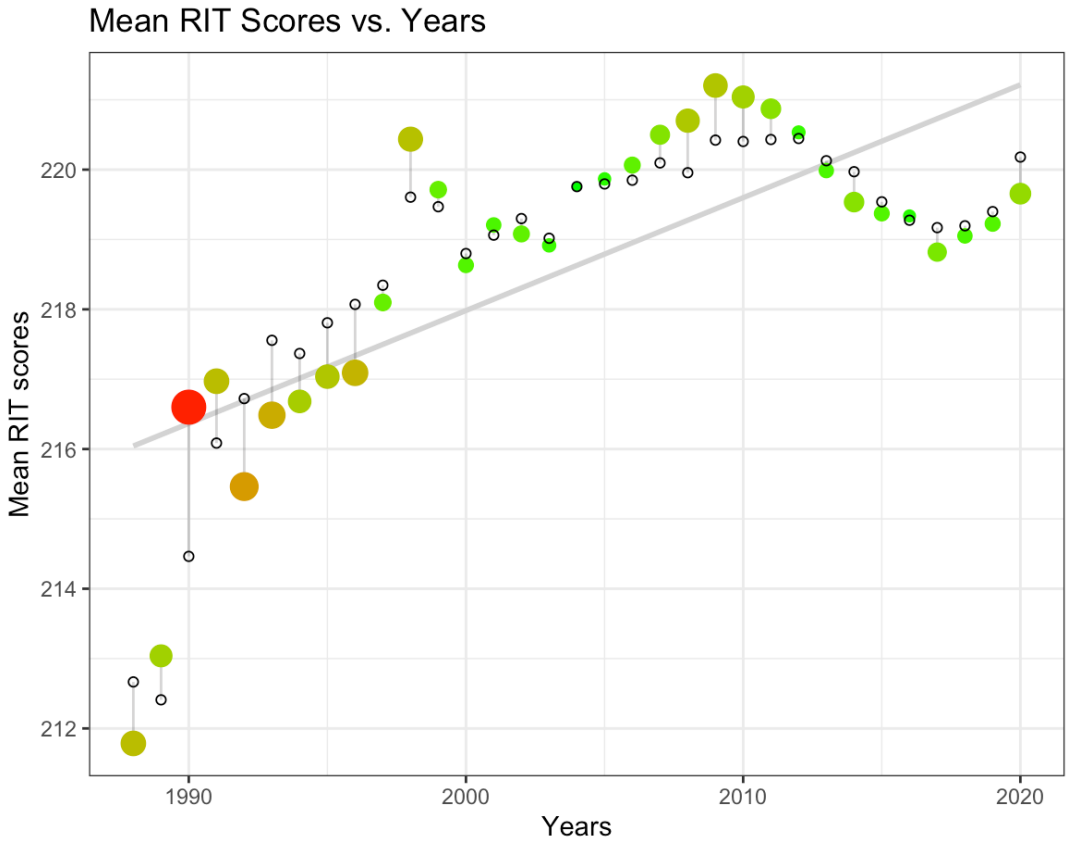
- **Data Grouped by Year, RIT (mean), Grade (mean), Test Levels (mean)**
- **Summary of the Linear Model**

```
Call:
lm(formula = mean_rit ~ mean_test_level + year + mean_grade,
    data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2596 -0.3502 -0.1023  0.4034  2.1365

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.67434   56.47504   0.012   0.9906
mean_test_level  3.72547    1.59000   2.343   0.0262 *
year             0.05036    0.02053   2.453   0.0204 *
mean_grade       8.56310    1.23028   6.960 1.19e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7162 on 29 degrees of freedom
Multiple R-squared:  0.9026,    Adjusted R-squared:  0.8925
F-statistic: 89.57 on 3 and 29 DF,  p-value: 9.085e-15
```

	year	mean_rit	mean_test_level	mean_grade	predicted	residuals
1	1988	211.7857	18.93608	4.827289	212.6659	-0.880143354
2	1989	213.0393	19.11607	4.713175	212.4096	0.629670310
3	1990	216.5998	19.26472	4.882444	214.4632	2.136544030
4	1991	216.9709	19.33097	5.037168	216.0853	0.885559637
5	1992	215.4626	19.38458	5.082335	216.7221	-1.259580814

Model Plots (with Regression Line):





Analysis of Linear Model (by Residual Values)*

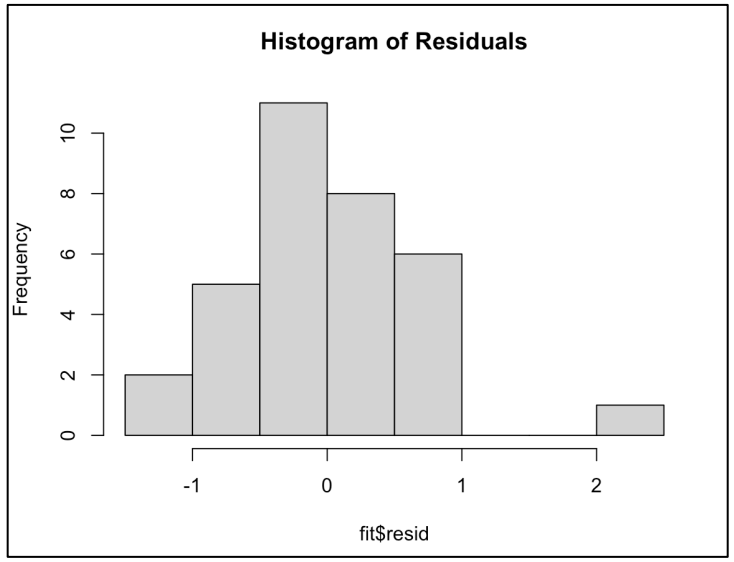
Criteria for Analysis:

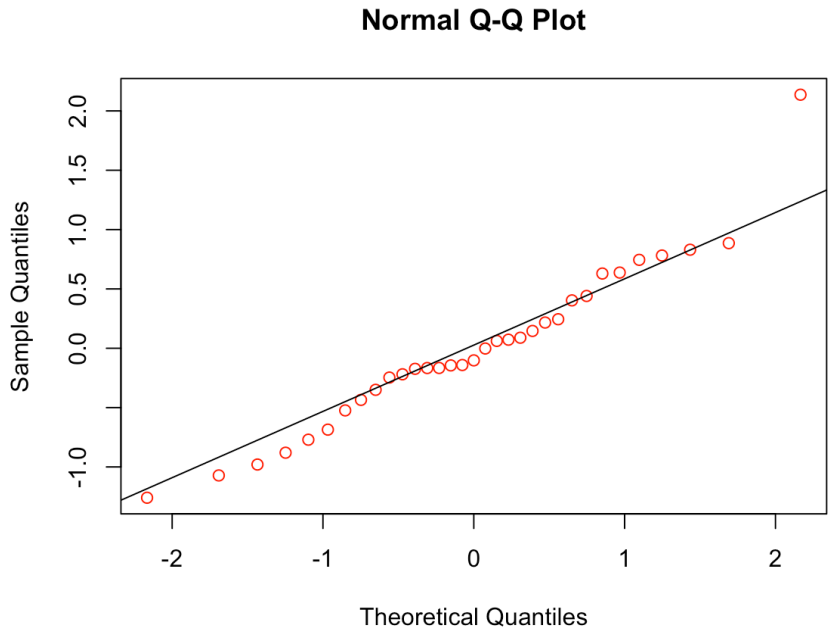
- The mean of residuals is 0.
- The distribution of residuals is normal.
- All the residuals are uncorrelated.
- Variance of residuals is Homoscedastic (constant).

* A residual is a vertical distance between Data Point (actual value) and Regression Line. Thus, residuals are equal to values in the data.

Test Normality of the residual distribution

- **Jarque – Bera Normality Test (To compare p-values with other models)**
- **Null Hypothesis for JB test: data is perfectly normalized (or normally distributed)**
- **Histogram**
- **Normal Q-Q plot**





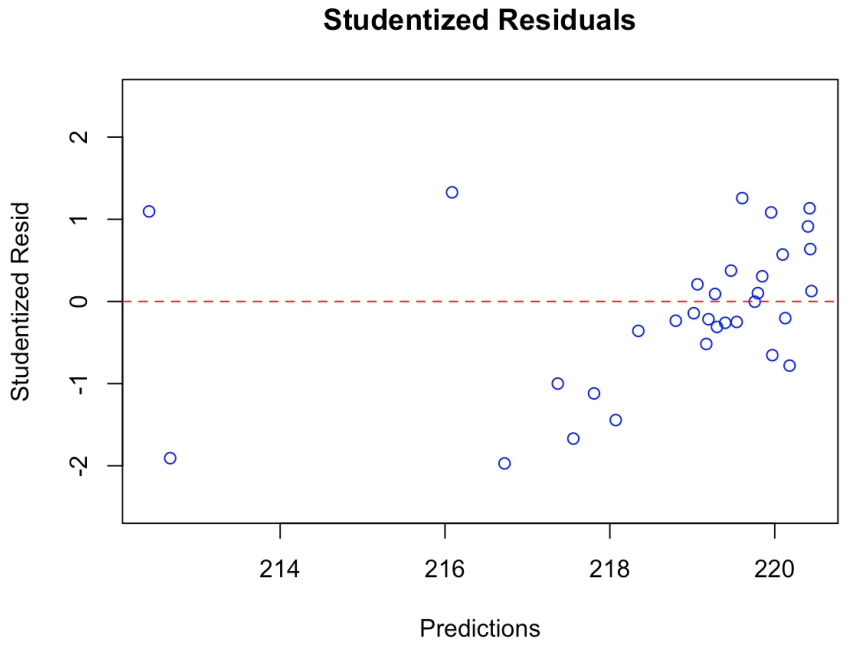

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 4.8919
  P VALUE:
    Asymptotic p Value: 0.08664
```

Observation on Normality:

- P-value is small (0.087) but much larger than in comparison with all other models that we've created before.
- Reject Null Hypothesis as Data is not perfectly normalized but still is somewhat normalized as shown in the histogram.
- Q-Q plot shows only one point off the normal line but others are relatively close to the normal line.
- It's clear from the histogram and Q-Q plot that residuals are normally distributed.

Test Variance of the residual distribution

- **Studentized Residuals**



Observation:

- Plot shows that there aren’t any obvious outliers. Based on the plot it's okay to assume the constant variance assumption.

Test Correlation between Residuals

- **Durbin – Watson independency Test (To compare p-values with other models)**
- **Null Hypothesis for DW test: residuals of time series data are completely uncorrelated**

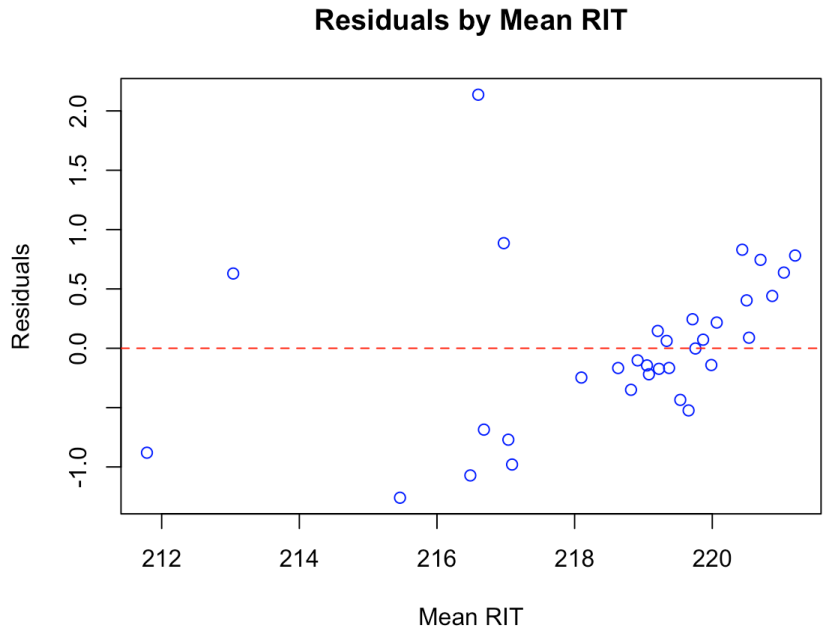
```
        Durbin-Watson test

data:  fit
DW = 0.9666, p-value = 0.0001096
alternative hypothesis: true autocorrelation is greater than 0
```

Observation:

- P-value is small (0.0001) but still much larger than in comparison with all other models that we've created before.
- Reject or Fail to reject Null Hypothesis ? there's some correlation but least correlation.
- We can conclude that the Variables we've used in this model has least correlation than any other combination of variables.

Test Mean of all Residuals

- **Plot of Residuals by RIT (mean)**



Observation:

- half of the residuals are above 0 and half below (shows that mean of all residuals is 0)
- Also, calculation by mean(df$residuals) gives 0.000....

Conclusions of Linear Regression Model

- Both Multiple and Adjusted coefficient of determination (R-square) values of 0.90 and 0.89 respectively shows that predicted values are closely fitted to regression line.
- The observations resulting from each criteria satisfies the error assumptions.
- This model have passed some of the most important criteria of Residual analysis.
- It's okay to say that this is pretty solid linear model and it's giving correct predictions.

Overview of Findings

- Students who took test in more recent years have done better (RIT scores increasing)
  - Student performance is correlated with year
  - Lang and reading RIT scores per grade are increasing more linearly over the years
  - Math RIT scores follow a quadratic trend
- RIT scores are proportional to grades
- Distribution of RIT + standard deviation scores isn't changing significantly
- % increase in the number of students taking the test by years?
- % increase in the Mean_RIT scores by years?

Conclusion

- Re normalization and redesign of the test are not necessary
  - Distribution of test scores is consistent
- Students are doing better
  - Different types of students are taking (different motives for home schooling)
  - Students and parents know better how to prepare for the test
  - Students are more accustom to the test
  - Students are more accustom to standardized testing in general