DSI 8 – PROJECT 1

EDA and Data Wrangling for
SAT and ACT scores, 2017 - 2018

By Group 3:
Justin Tan, Chang Chu Hua, Alfred Tang
DSI 8, General Assembly
23 May 2019

INTRODUCTION

- SAT format changed in 2016
- Now there are 2 years of data to compare, post-change
- Task: Analyse data and present findings to College Board staff

PROBLEM STATEMENT

- In which states are participation rates increasing after re-design of SAT?

- Are more students achieving the benchmark scores after the re-design?

DATA CLEANING (2017)

- Issues:
  - Row counts not matching
    - ACT 2017 National Composite
  - Typos
    - SAT 2017 Maryland Math Score → 52 instead of 524

DATA CLEANING (2017)

- Issues (Cont'd):
  - Data Entry
    - ACT 2017 Wyoming Composite → 20.2x
    - ACT 2017 Maryland Science → 2.3 → 23.2
  - Data Types
    - Participation entered as strings
      - Replaced % with blank; converted to float

DATA GATHERING (2018)

- Issues:
  - Source Data not complete
    - Composite column was the only usable column
    - Separately gathered score data
  - Data not matching
    - SAT scores did not match Totals
    - +1/-1 differences ignored.
      - Difference is due to rounding
      - Officially stated by College Board.

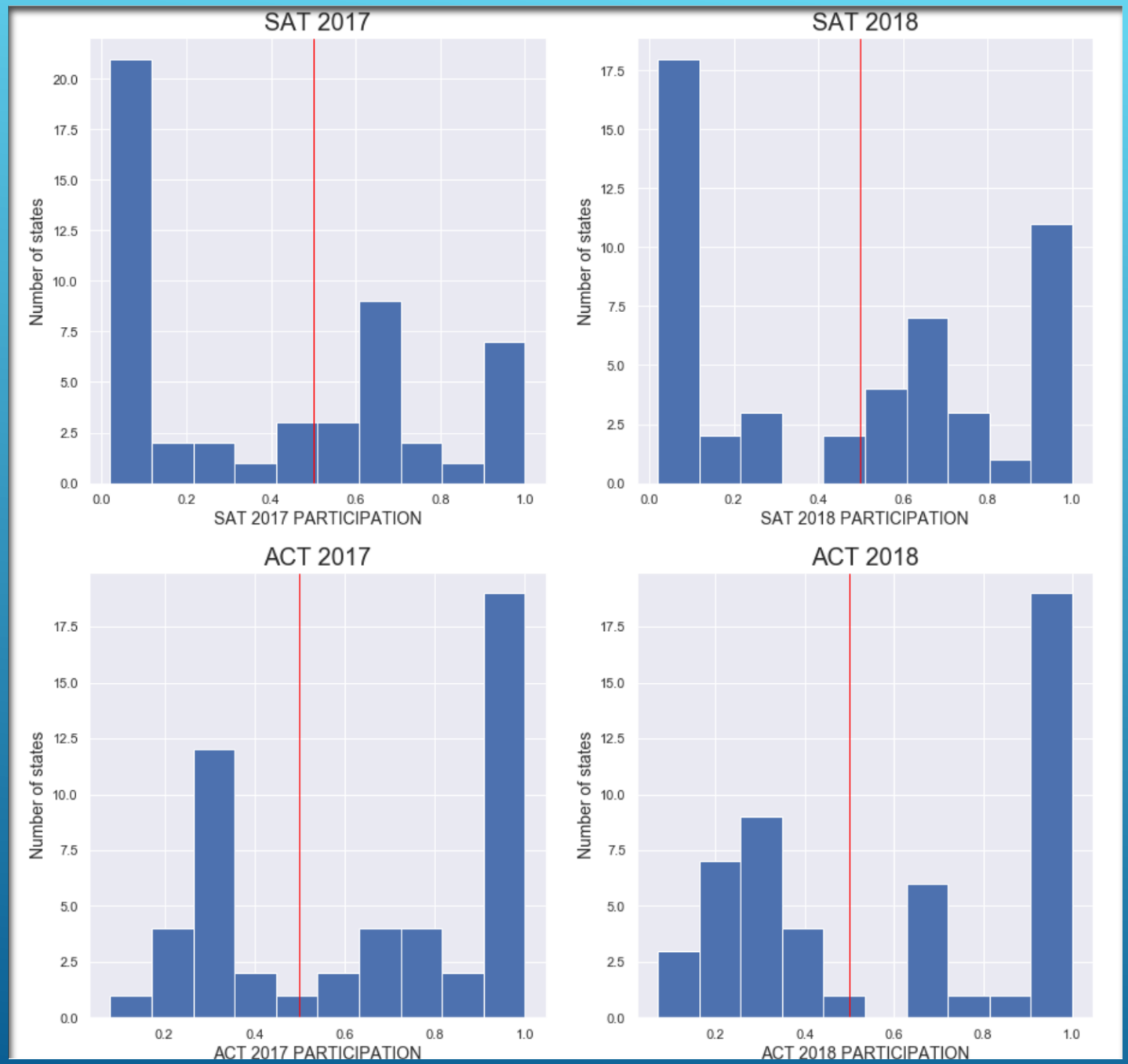
DATA CLEANING (2018)

- Issues:
  - Row counts not matching
    - 2018 National averages
  - Data Types
    - Participation again entered as strings

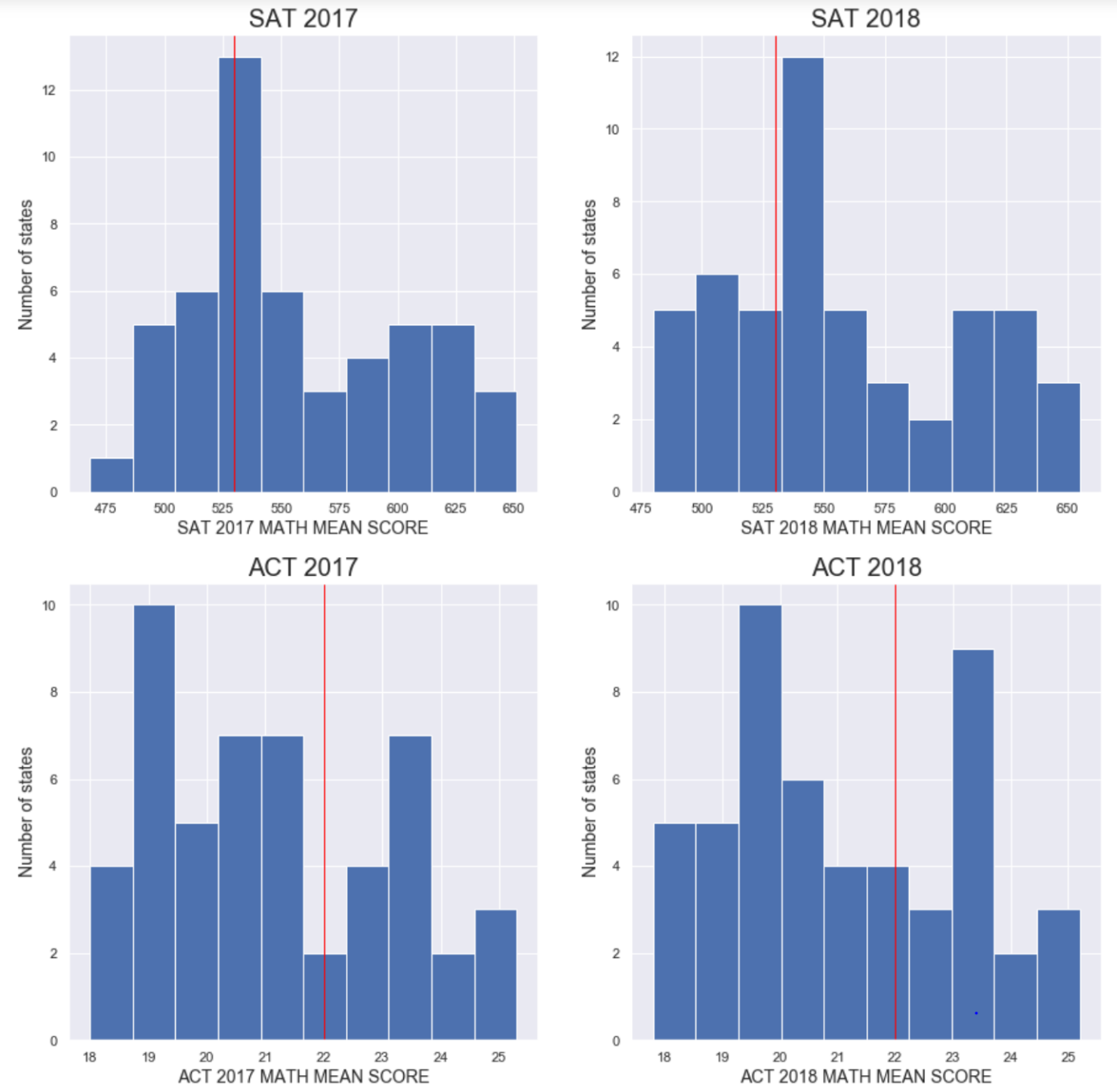
EDA: SUMMARY STATISTICS

	count	mean	std	min	25%	50%	75%	max
sat_2017_participation	51.0	0.398039	0.352766	0.02	0.040	0.38	0.660	1.0
sat_2017_erw	51.0	569.117647	45.666901	482.00	533.500	559.00	613.000	644.0
sat_2017_math	51.0	556.882353	47.121395	468.00	523.500	548.00	599.000	651.0
sat_2017_total	51.0	1126.098039	92.494812	950.00	1055.500	1107.00	1212.000	1295.0
act_2017_participation	51.0	0.652549	0.321408	0.08	0.310	0.69	1.000	1.0
act_2017_english	51.0	20.931373	2.353677	16.30	19.000	20.70	23.300	25.5
act_2017_math	51.0	21.182353	1.981989	18.00	19.400	20.90	23.100	25.3
act_2017_reading	51.0	22.013725	2.067271	18.10	20.450	21.80	24.150	26.0
act_2017_science	51.0	21.450980	1.739353	18.20	19.950	21.30	23.200	24.9
act_2017_composite	51.0	21.519608	2.020695	17.80	19.800	21.40	23.600	25.5
sat_2018_participation	51.0	0.466275	0.380142	0.02	0.045	0.52	0.795	1.0
sat_2018_erw	51.0	567.294118	45.317676	497.00	535.000	552.00	616.500	643.0
sat_2018_math	51.0	557.254902	48.887562	480.00	521.500	547.00	600.500	655.0
sat_2018_total	51.0	1124.666667	93.867069	977.00	1062.500	1099.00	1220.000	1298.0
act_2018_participation	51.0	0.616471	0.340810	0.07	0.285	0.66	1.000	1.0
act_2018_english	51.0	20.988235	2.446356	16.60	19.100	20.20	23.700	26.0
act_2018_math	51.0	21.125490	2.035765	17.80	19.400	20.70	23.150	25.2
act_2018_reading	51.0	22.015686	2.167245	18.00	20.450	21.60	24.100	26.1
act_2018_science	51.0	21.345098	1.870114	17.90	19.850	21.10	23.050	24.9
act_2018_composite	51.0	21.486275	2.106278	17.70	19.950	21.30	23.550	25.6

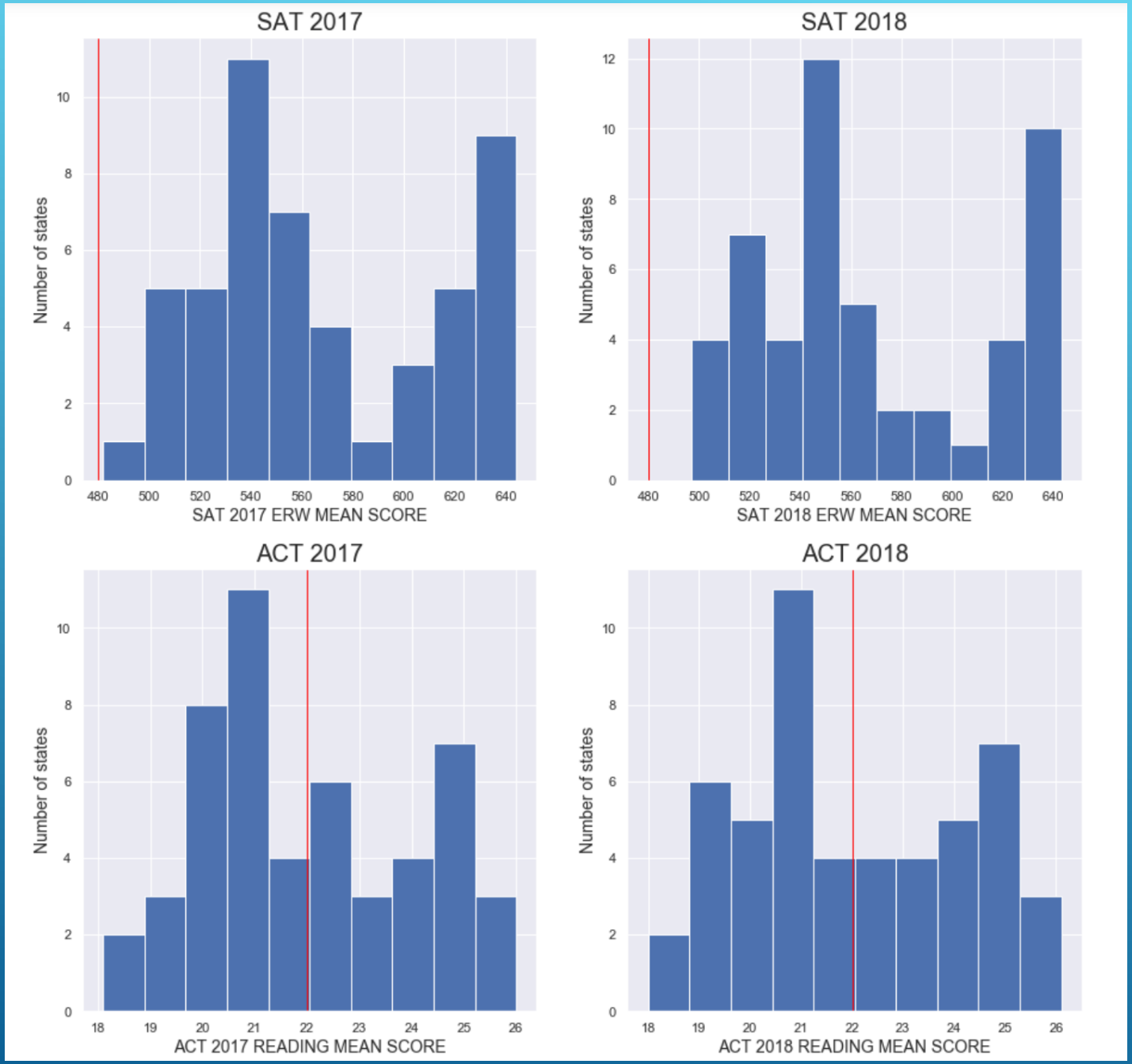
Participation



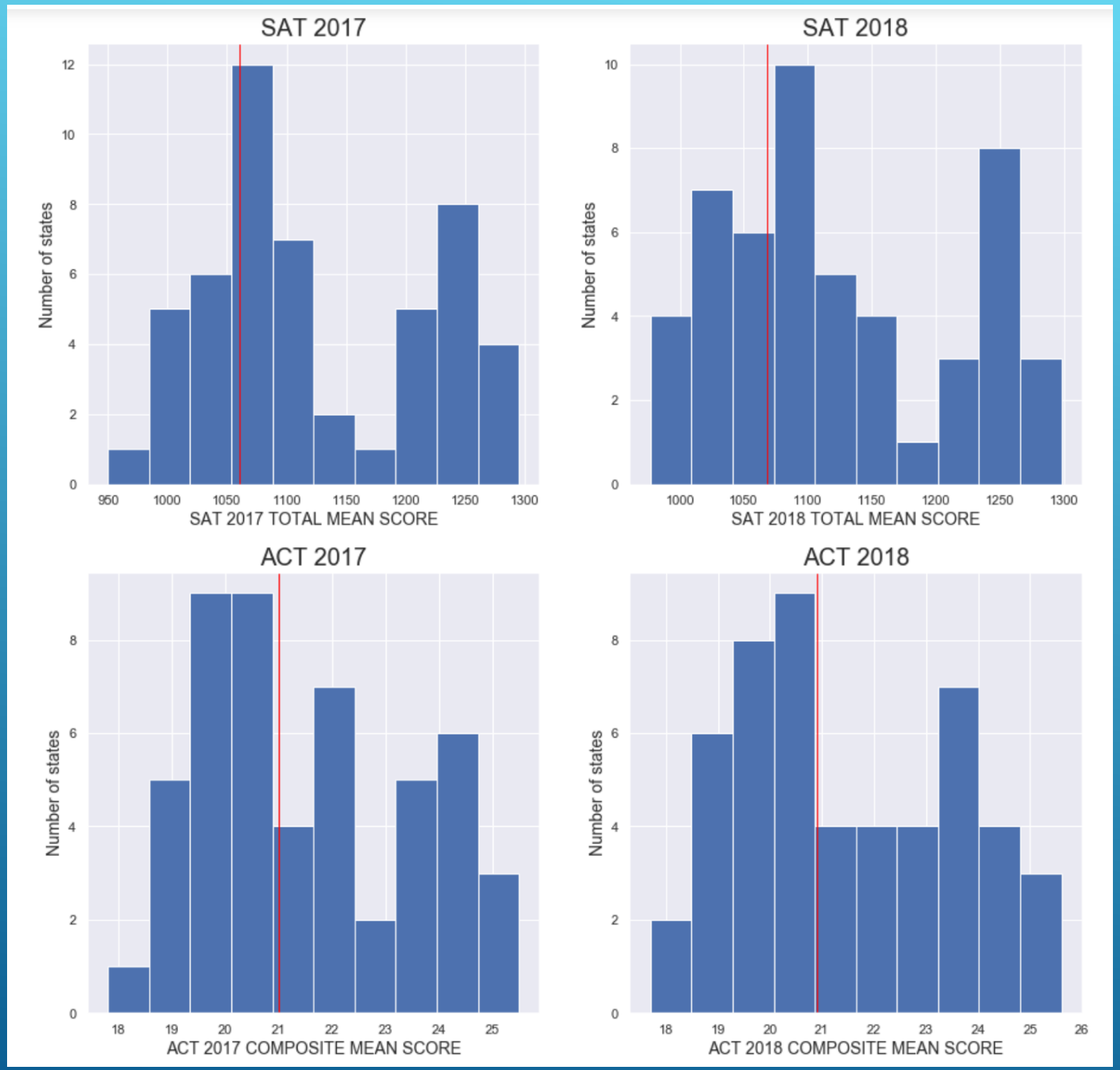
Mathematics



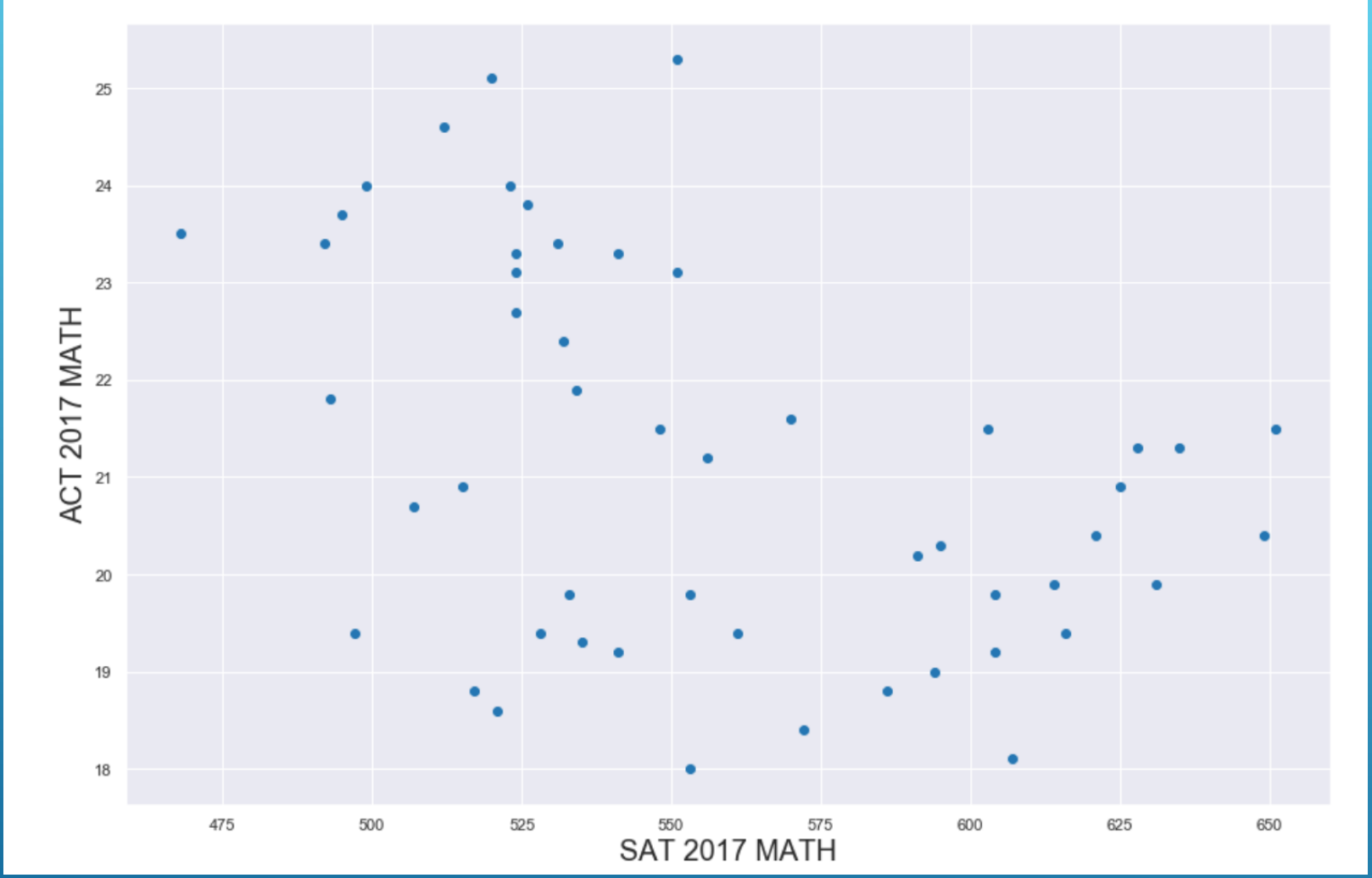
ERW vs Reading



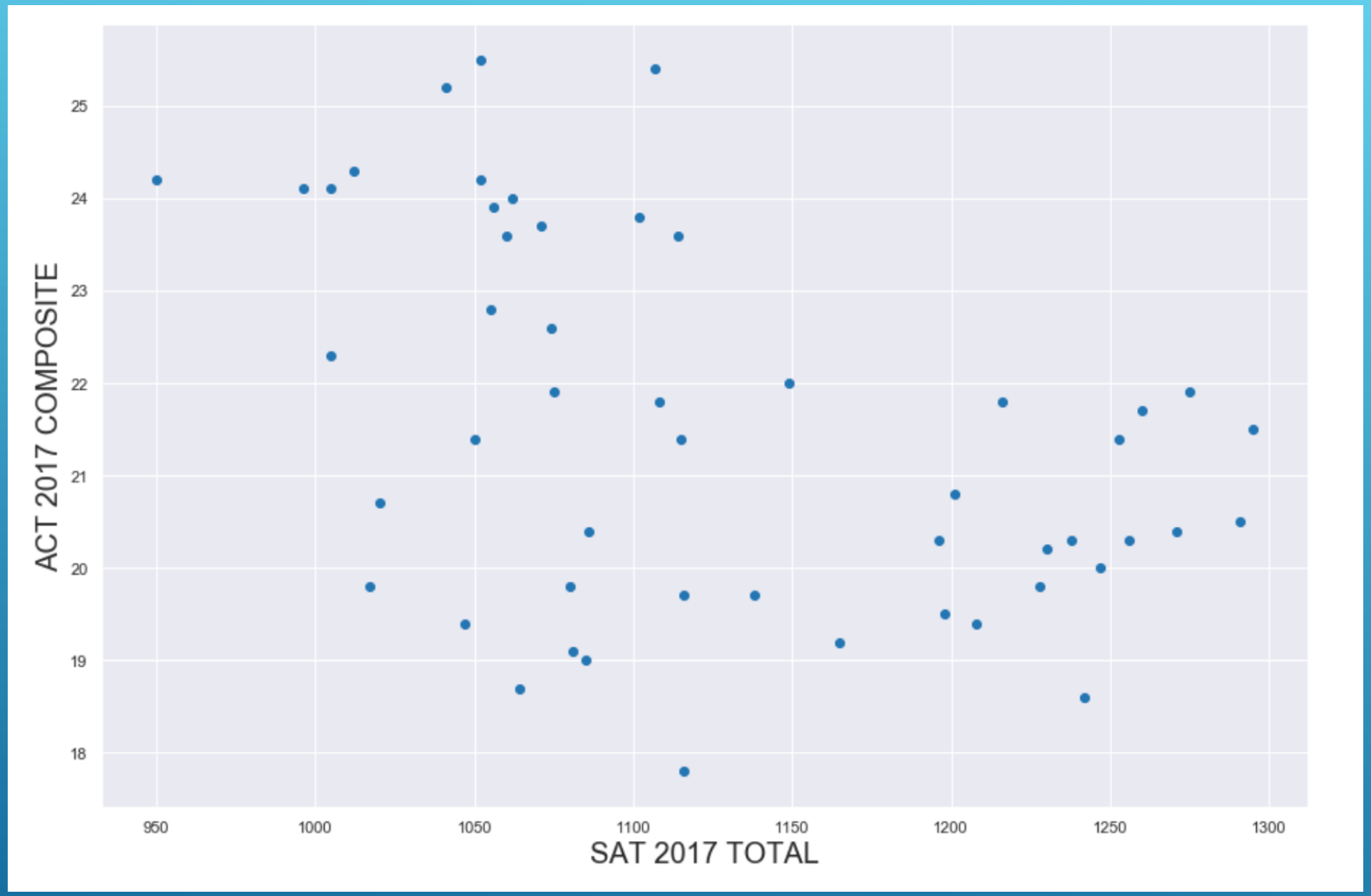
Total Mean vs Composite Mean

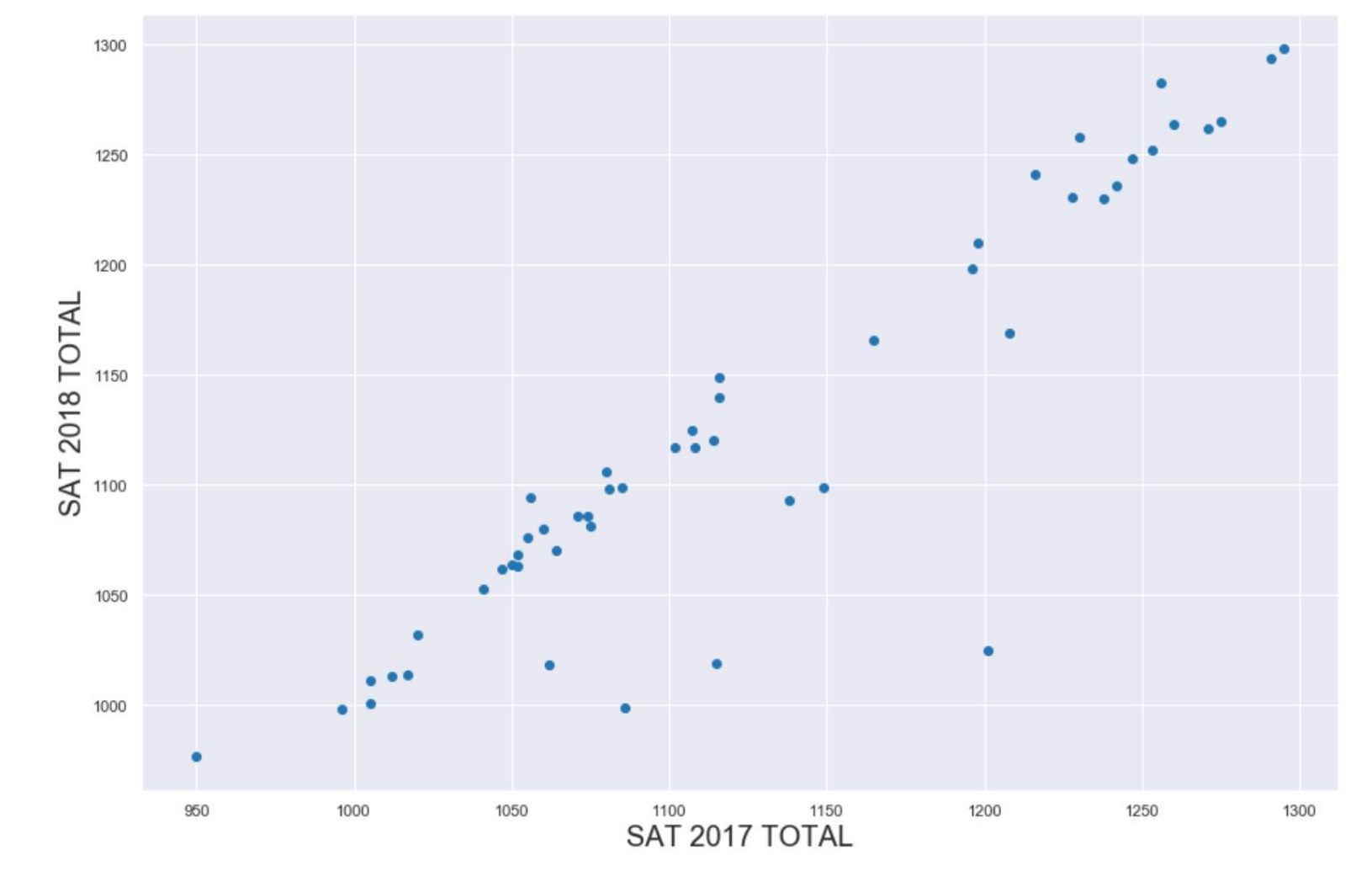


ACT Math vs SAT Math 2017

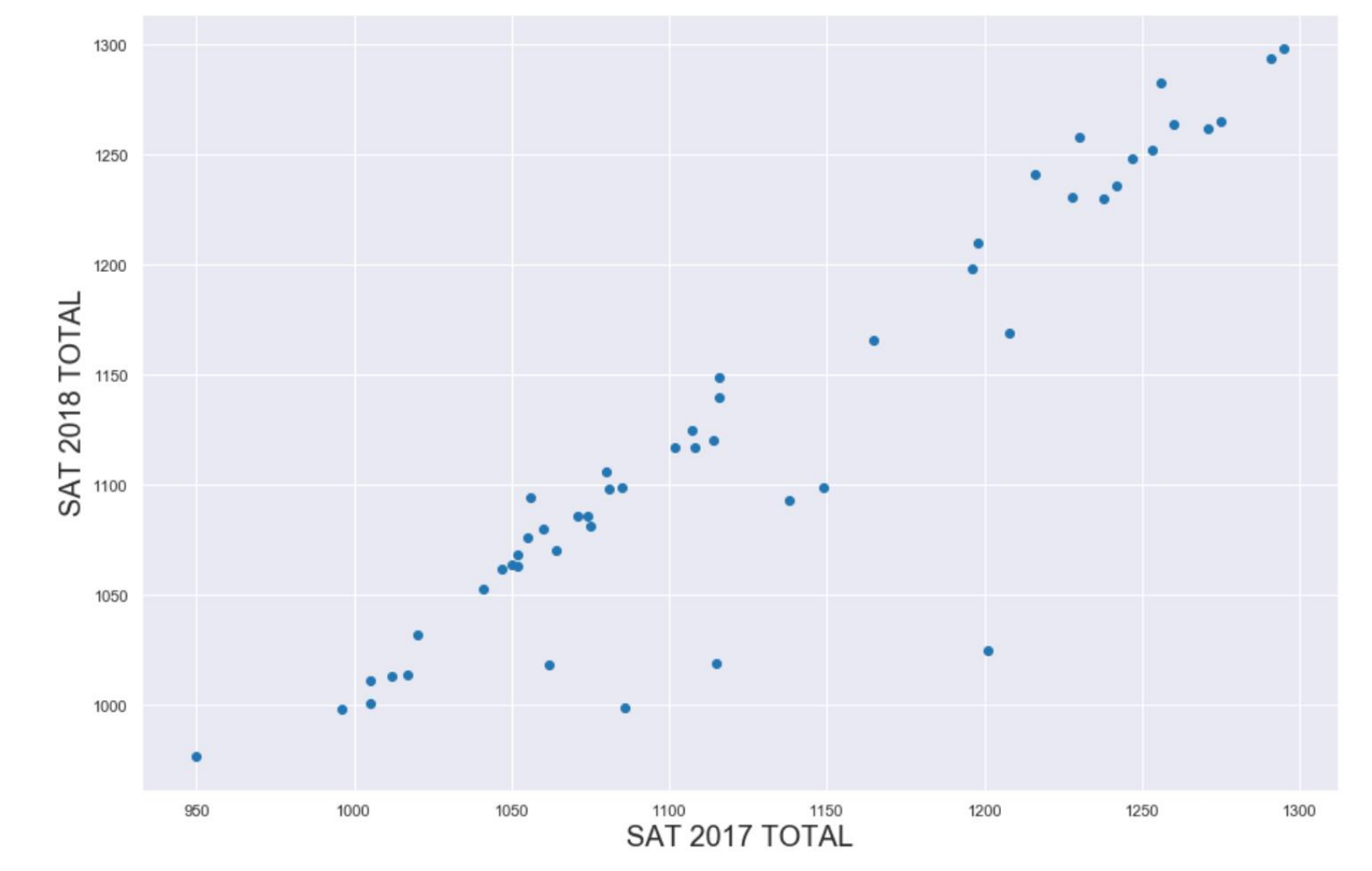


ACT 2017 Comp Vs SAT 2017 Total

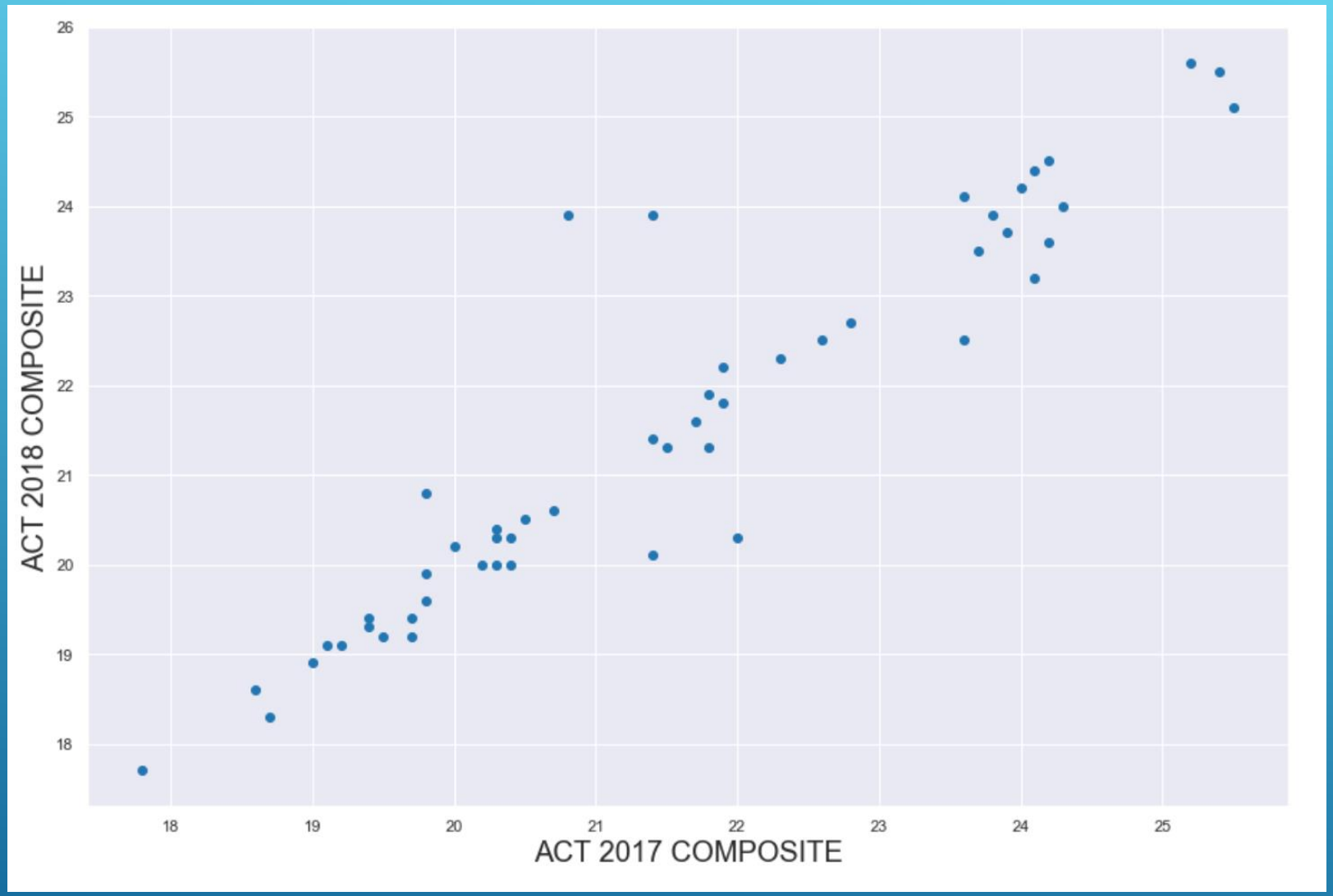




SAT 2018 Total Vs SAT 2017 Total

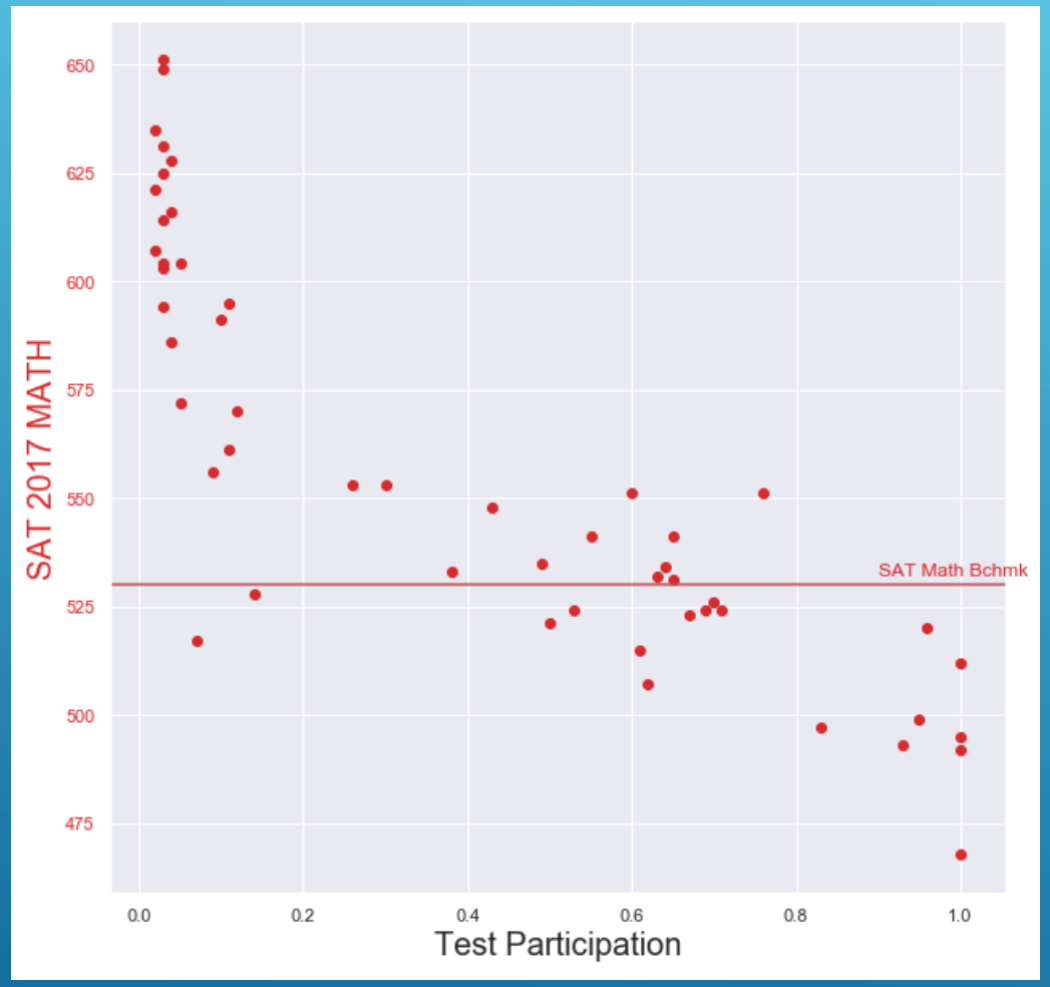


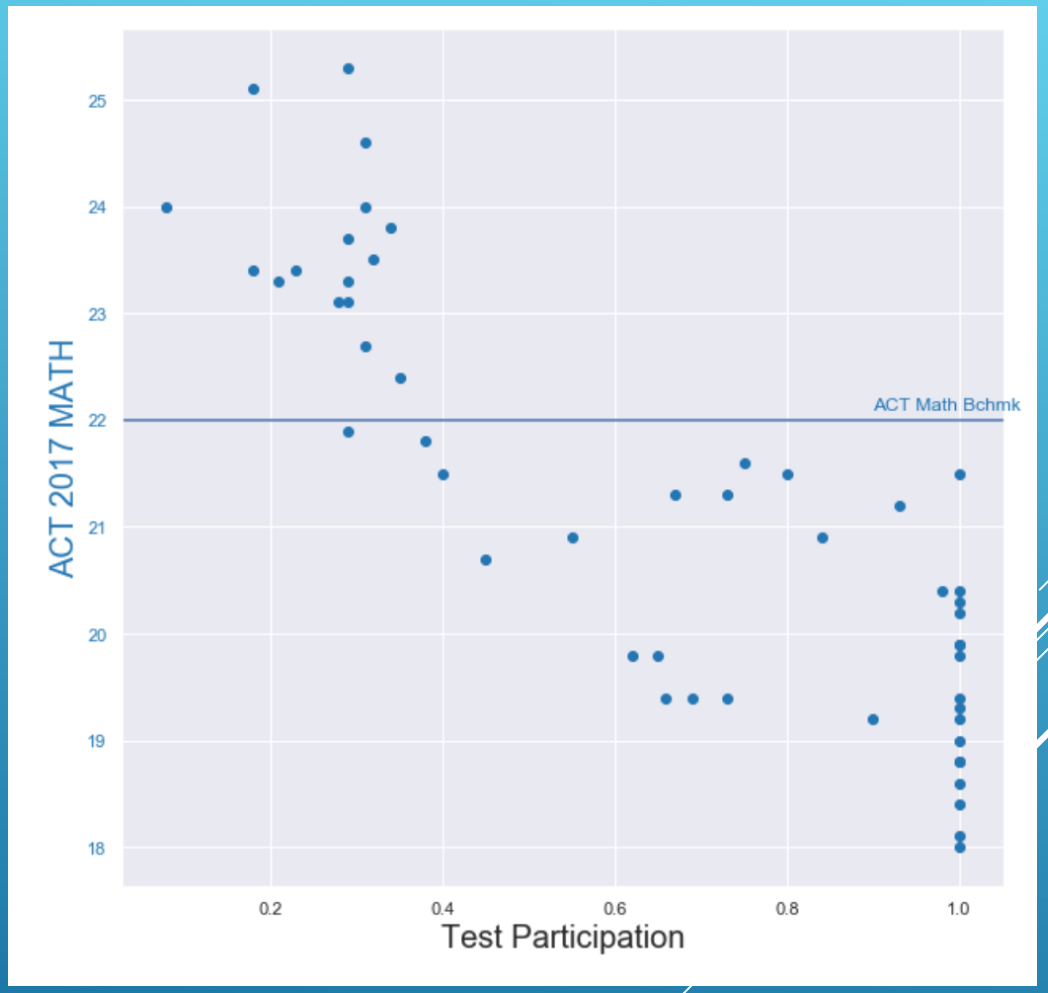
SAT 2018 Total Vs SAT 2017 Total



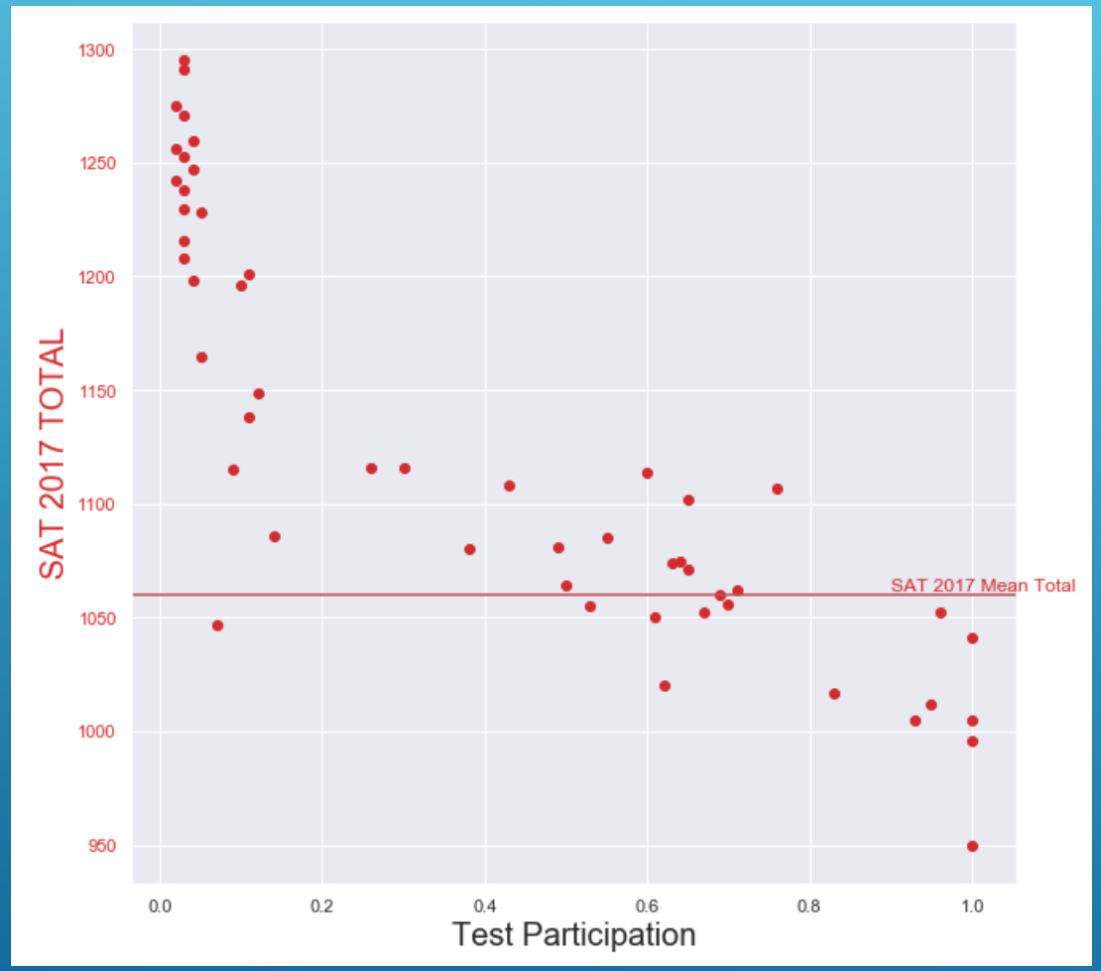
ACT 2018 Comp Vs ACT 2017 Comp

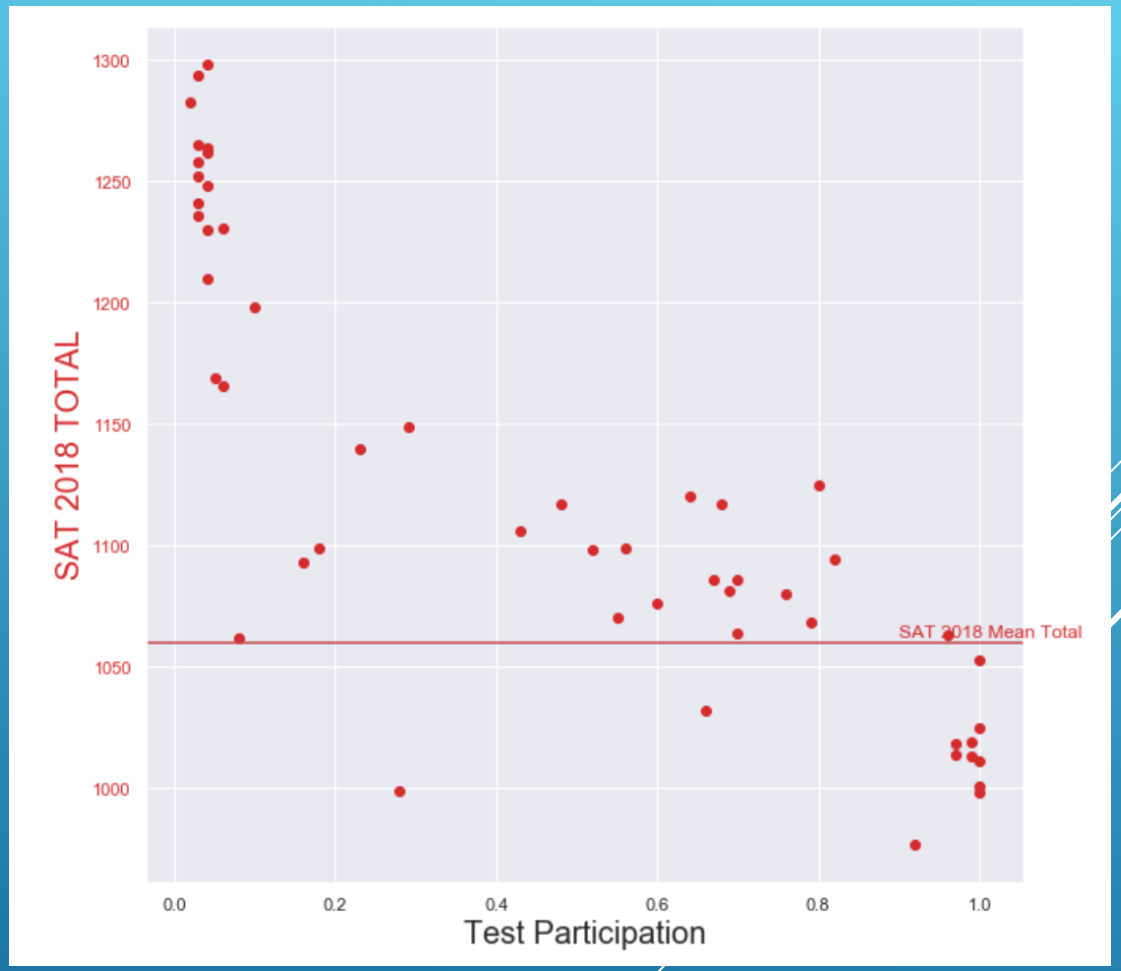
SAT & ACT MATH 2017 VS BENCHMARKS



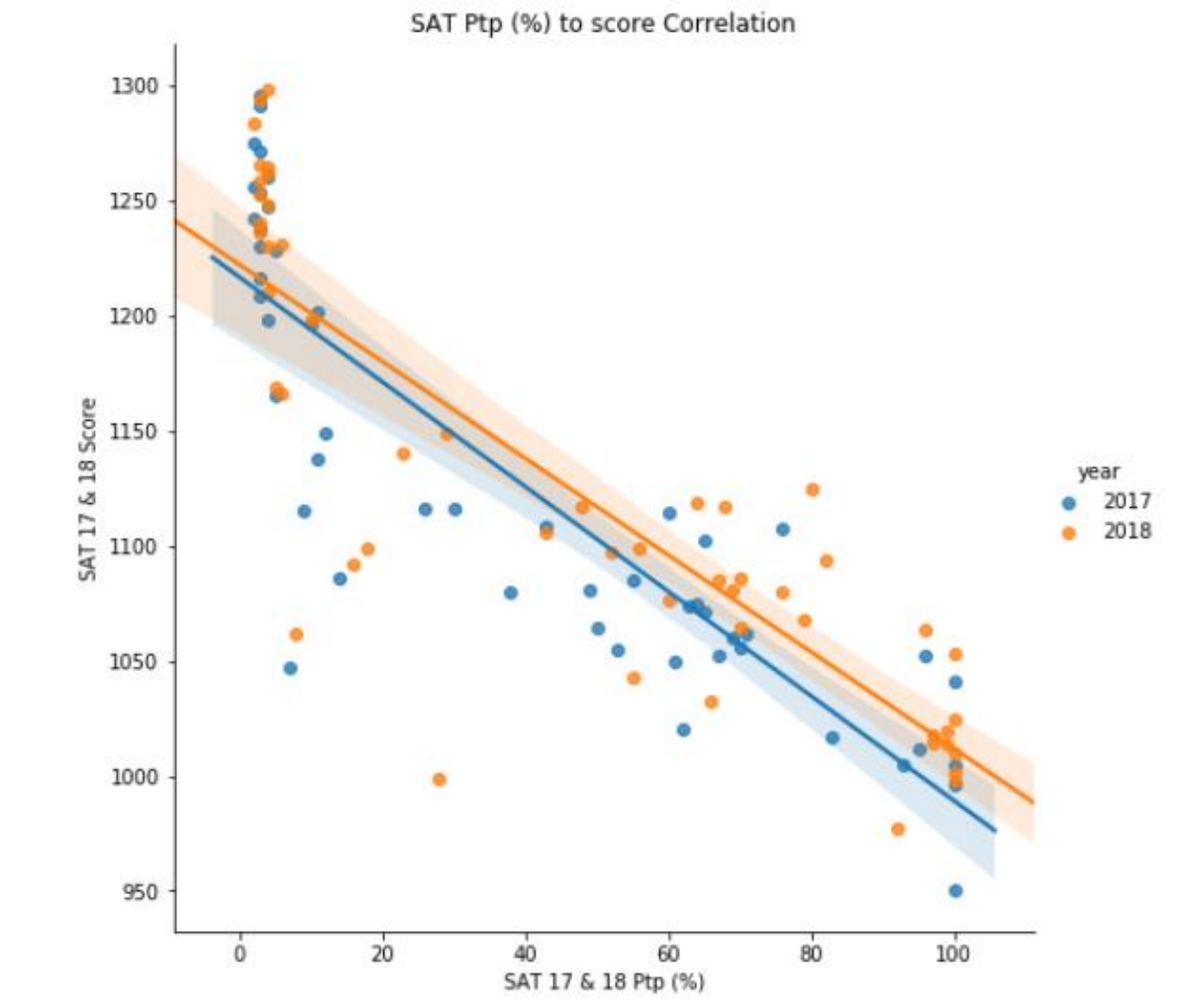


SAT TOTAL 2017 – 2018 VS BENCHMARKS





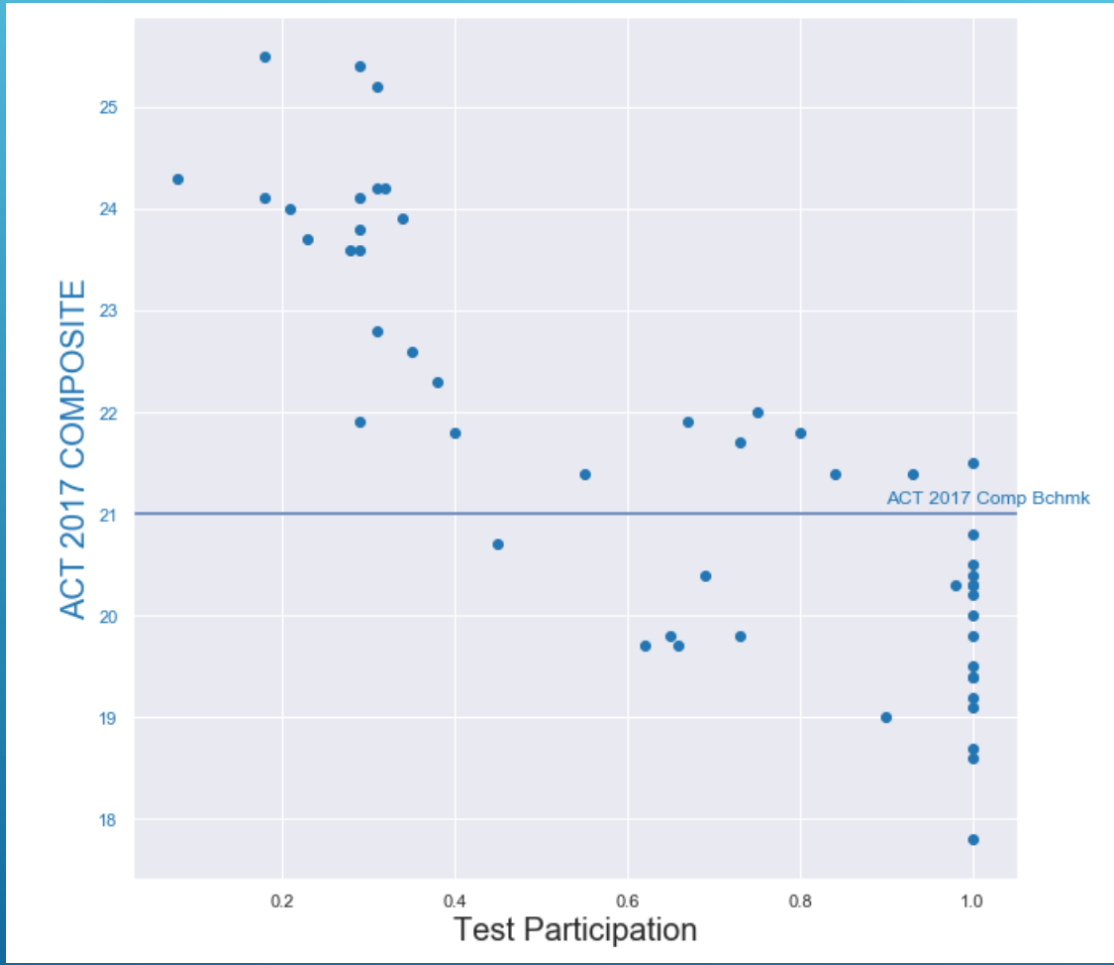
SAT TOTAL 2017 & 2018 VS PARTICIPATION

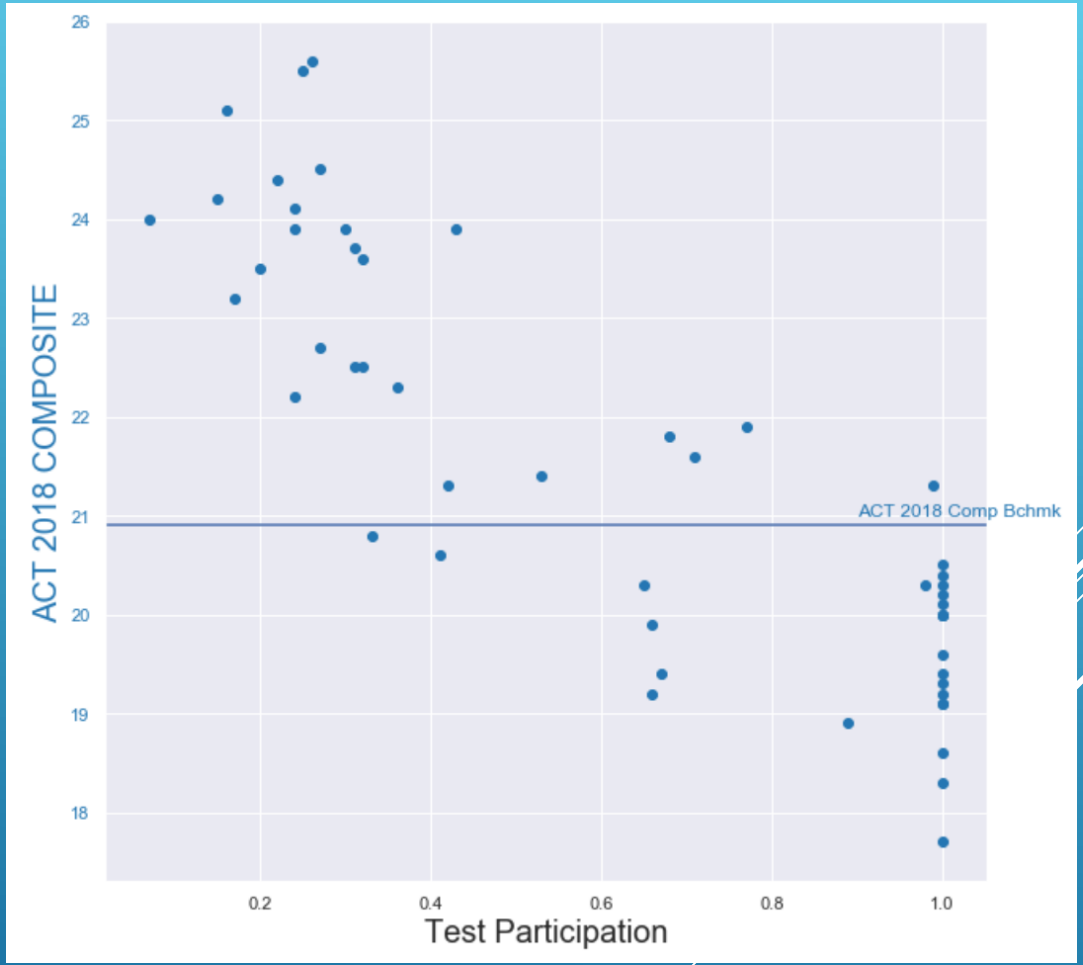


Correlation is flattening

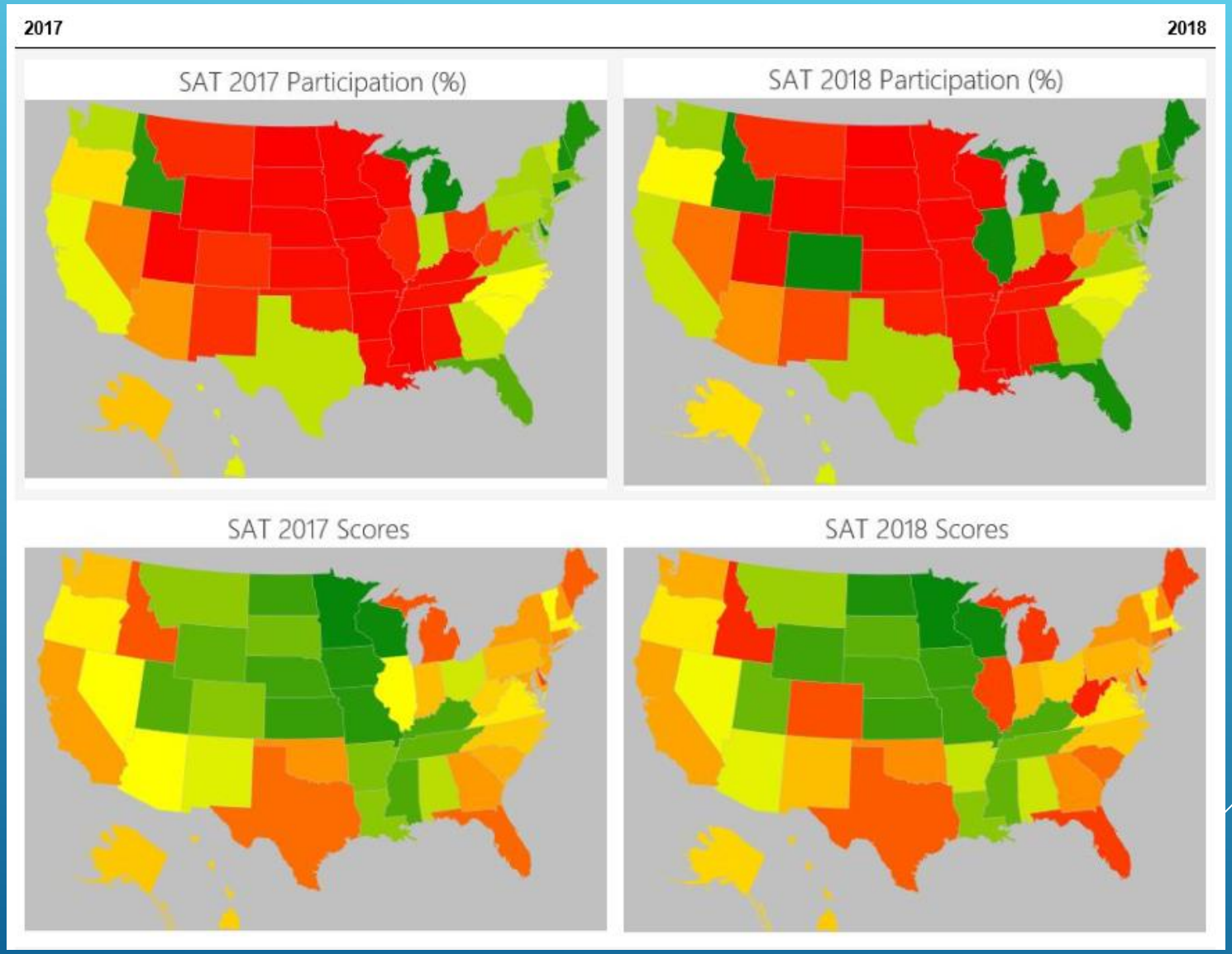
This suggests the change in SAT test structure is normalising results

ACT TOTAL 2017 – 2018 VS BENCHMARKS

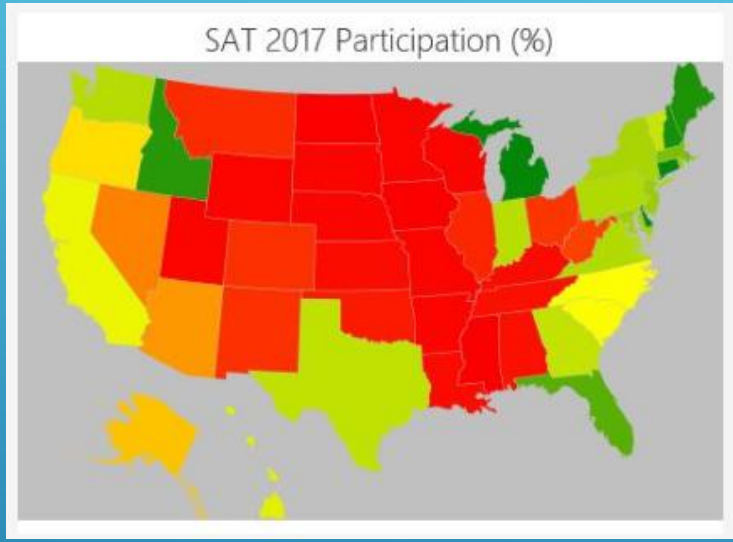


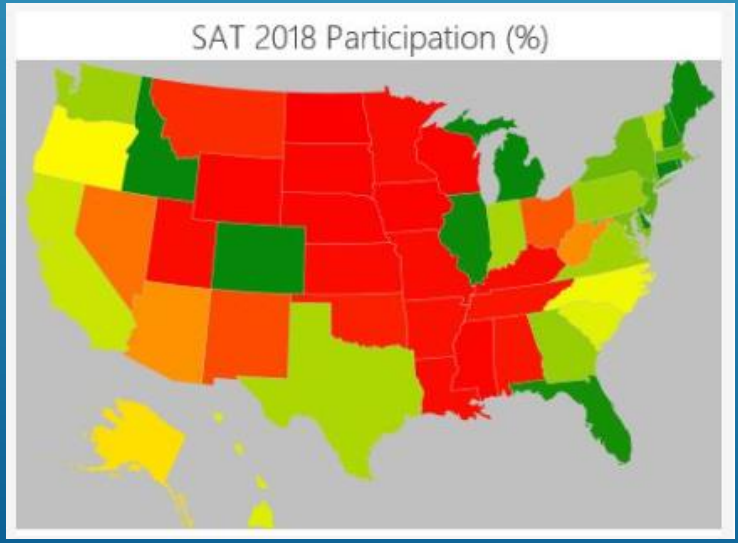


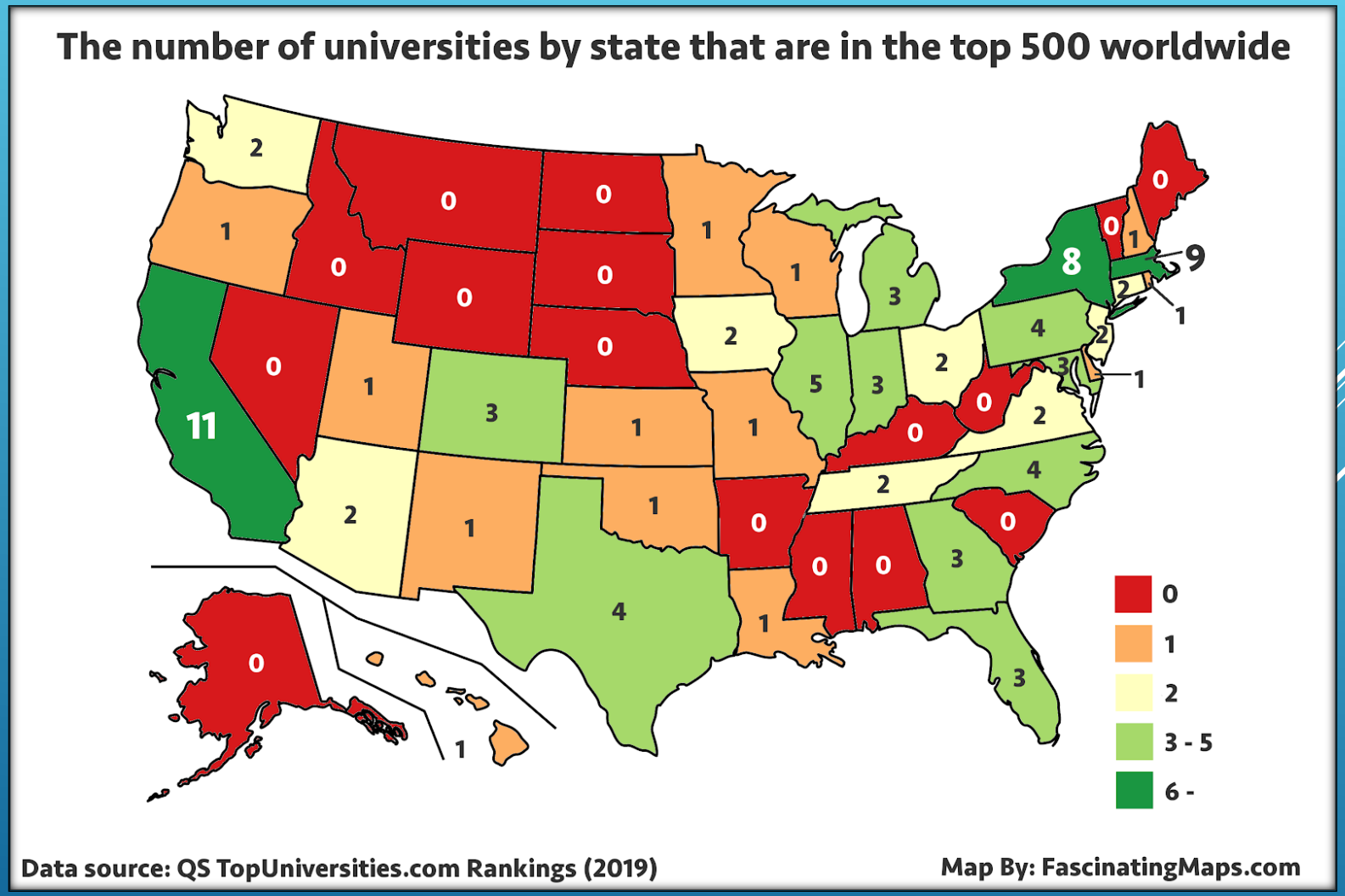
SAT TOTAL 2017 & 2018 VS PARTICIPATION



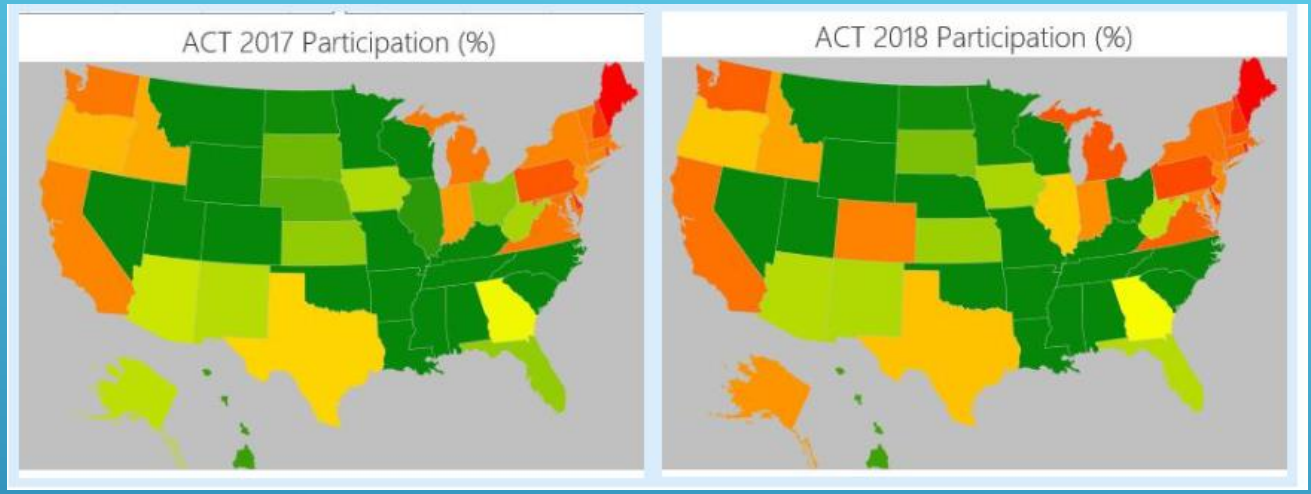
SAT PARTICIPATION VS TOP US UNIS



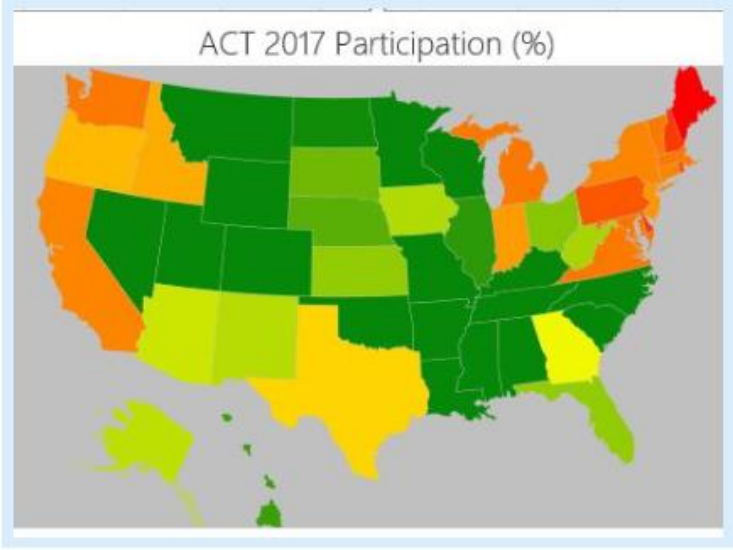


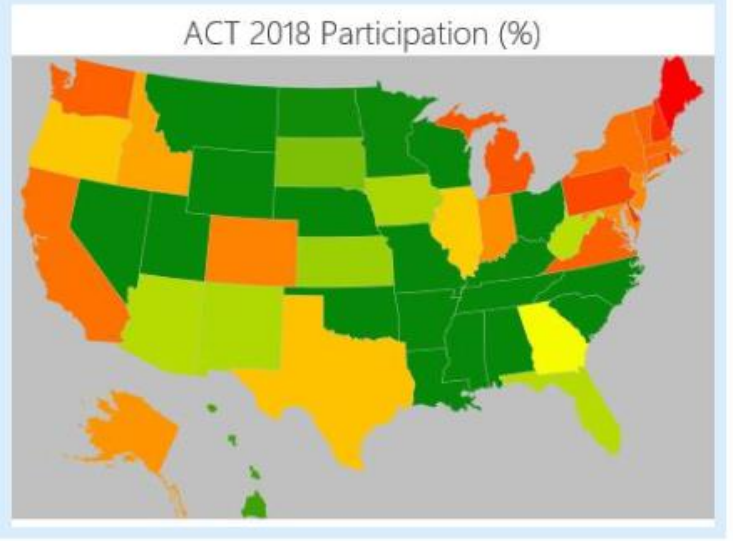


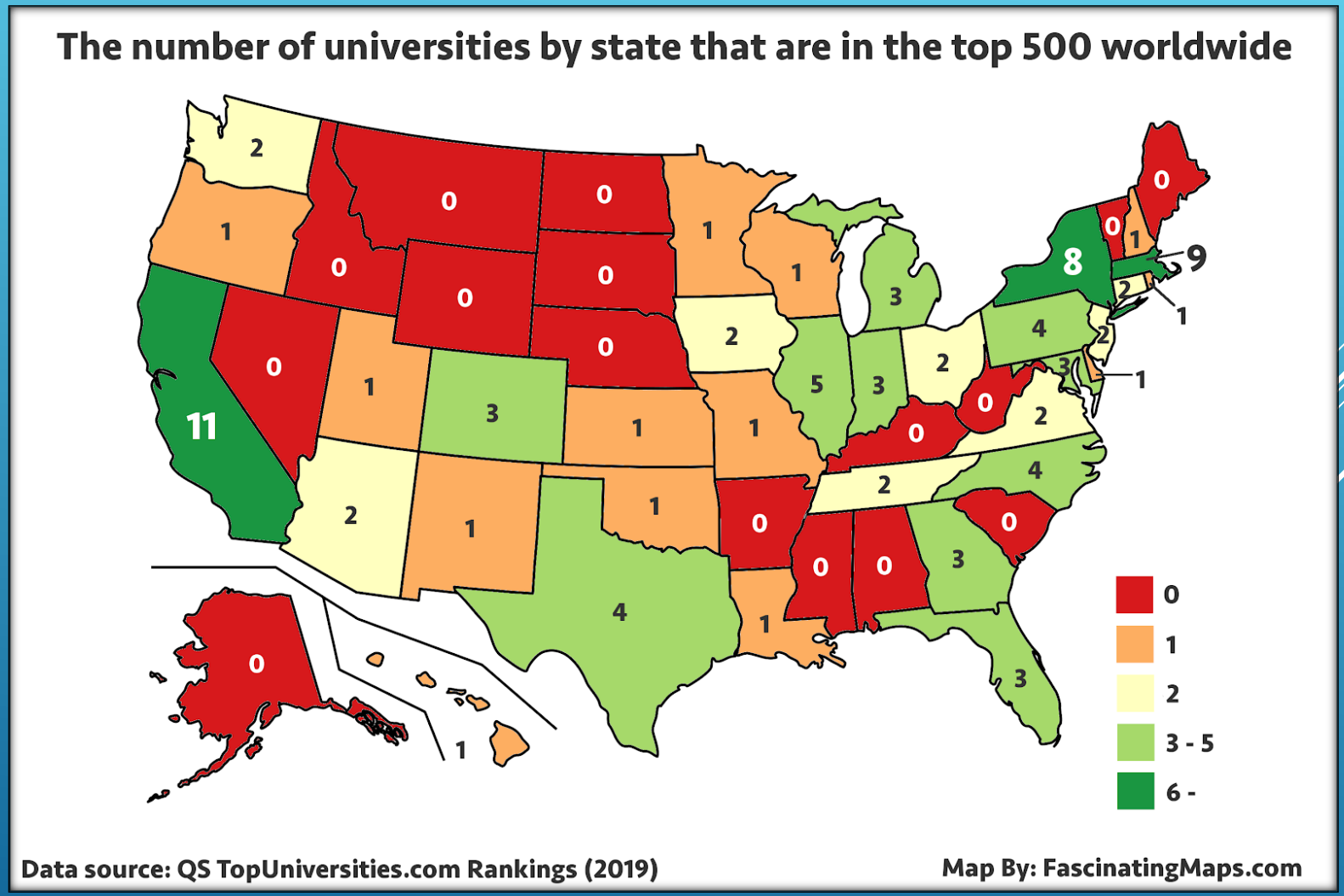
ACT TOTAL 2017 – 2018 VS BENCHMARKS



SAT TOTAL 2017 & 2018 VS PARTICIPATION







SCORES VS TEST PARTICIPATION

- **Benchmarks**
  - Many states do not meet the ACT benchmarks or SAT Benchmarks for each test.
  - Most state mean SAT Total scores are at or above the national mean.
  - Approximately half of all ACT state mean composite scores are at or above the national benchmark.
  - As participation levels increase, mean Total and Composite scores decrease.

SCORES VS TEST PARTICIPATION

- **Participation**
  - Numerous states have full participation in the ACTs and SATs.
  - More "normal" results are included where all students participate in the tests, thus lowering the state averages
  - Most states with high mean scores have low participation rates.

RELATIONSHIP: SAT & ACT PARTICIPATION

- It could be a justifiable comparison because:
  - Same kind of metric
    - Participation for test used in college admissions
  - Exactly the same scale (0.00 to 1.00)
    - It is possible to establish a reasonably accurate correlation value.

RELATIONSHIP: SAT & ACT PARTICIPATION

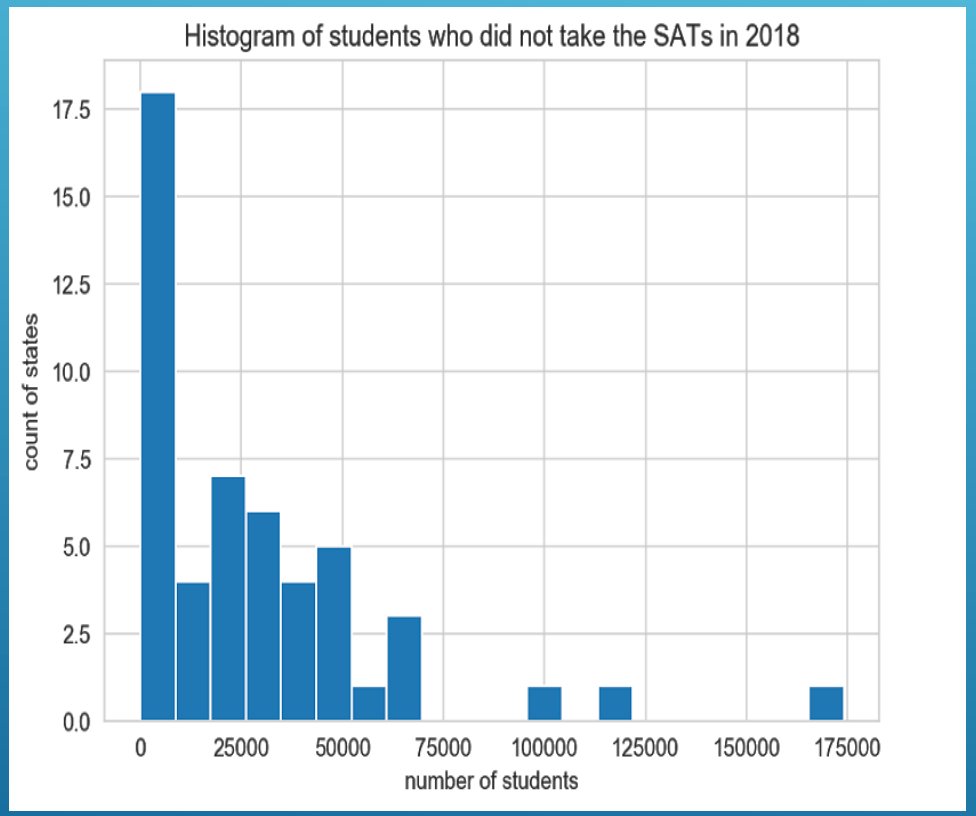
- However, without additional data, further issues can crop up.

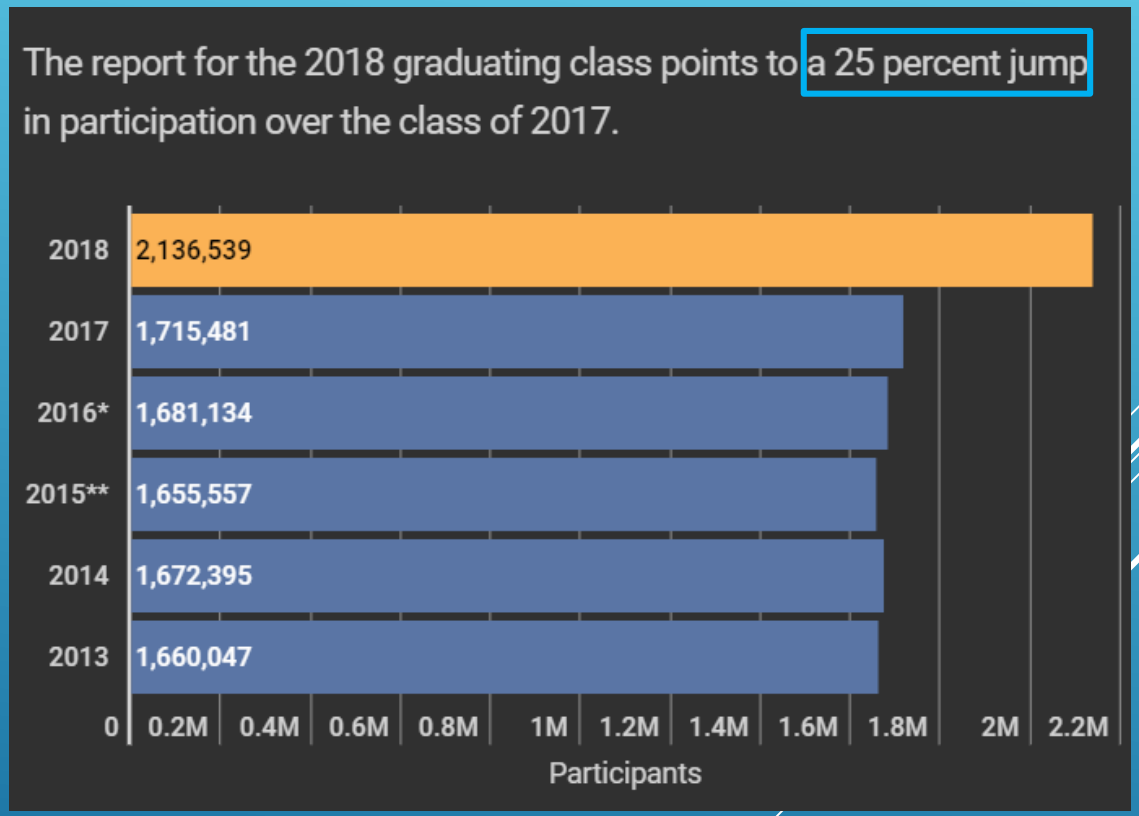
- Issues:
  - They are not entirely independent variables.
    - Test participation rules and incentives vary across the states
    - Geography influences test participation
  - Each is a major exam that can be used for college admission
    - Taking one test likely means not taking the other
    - SATs are more recognised for university admissions worldwide

RELATIONSHIP: SAT & ACT PARTICIPATION

- Issues (Cont'd):
  - Populations of each state are not reflected alongside participation data.
    - Some states, such as Florida, have high populations (147,058 students).
    - High population more likely to drastically lower mean scores than for a less populous state
      - 147,058 students in Florida vs. 14,834 students in Rhode Island
  - SAT and ACT participation have a strong negative correlation with each other (approx. -0.8)

RELATIONSHIP: SAT & ACT PARTICIPATION



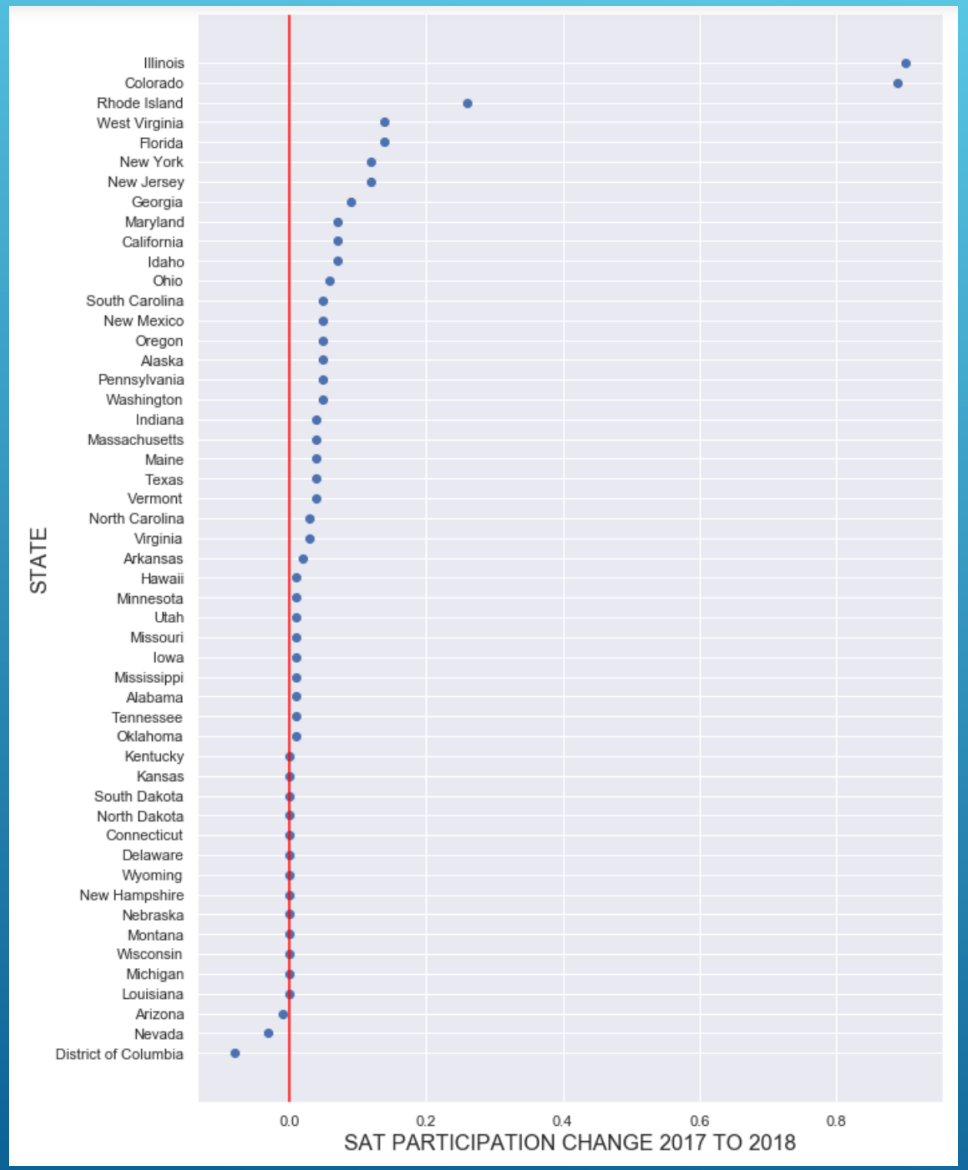


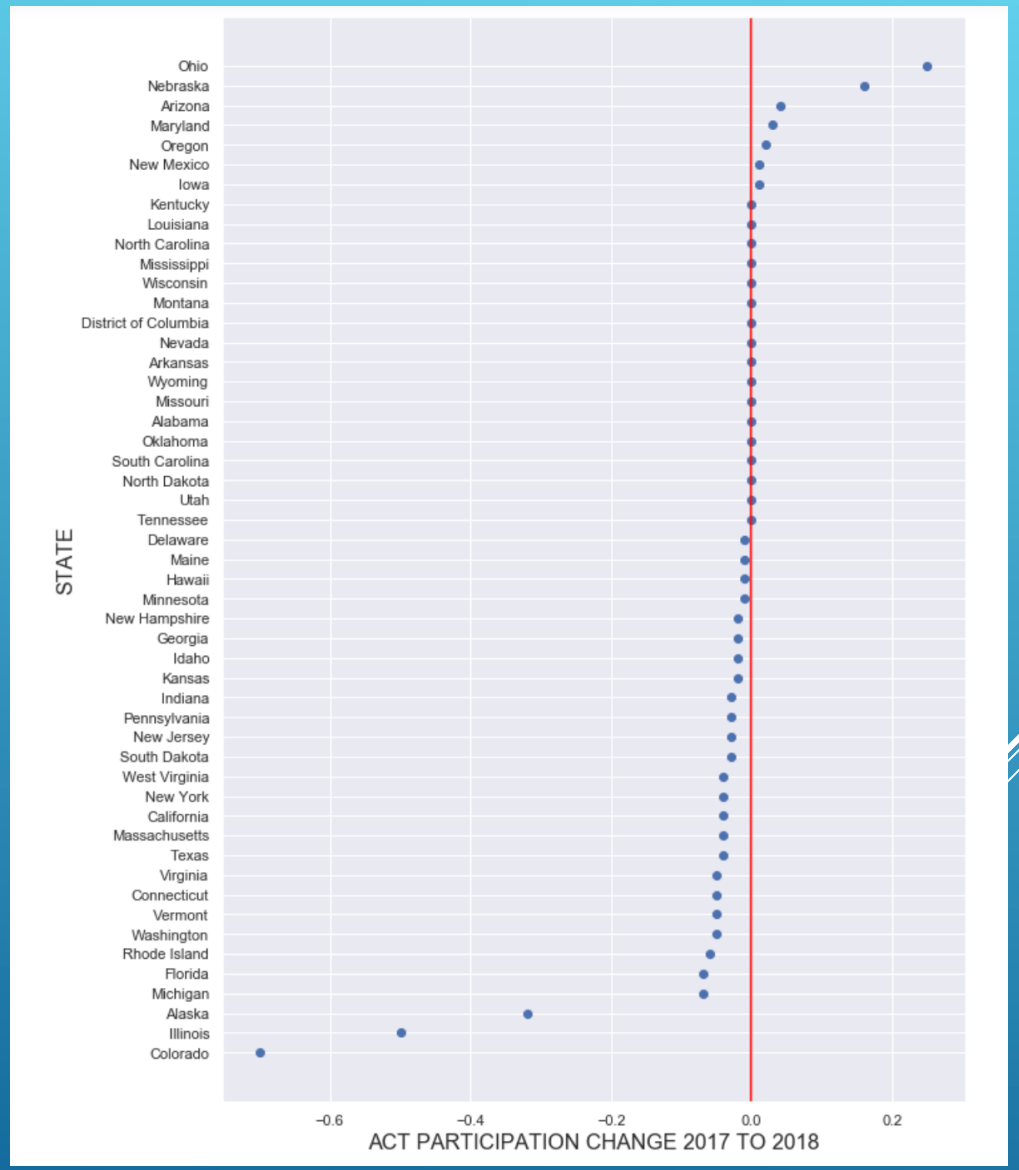
Source: https://www.edweek.org/ew/articles/2018/10/31/sat-scores-rise-as-number-of-test-takers.html

RELATIONSHIP: SAT & ACT PARTICIPATION

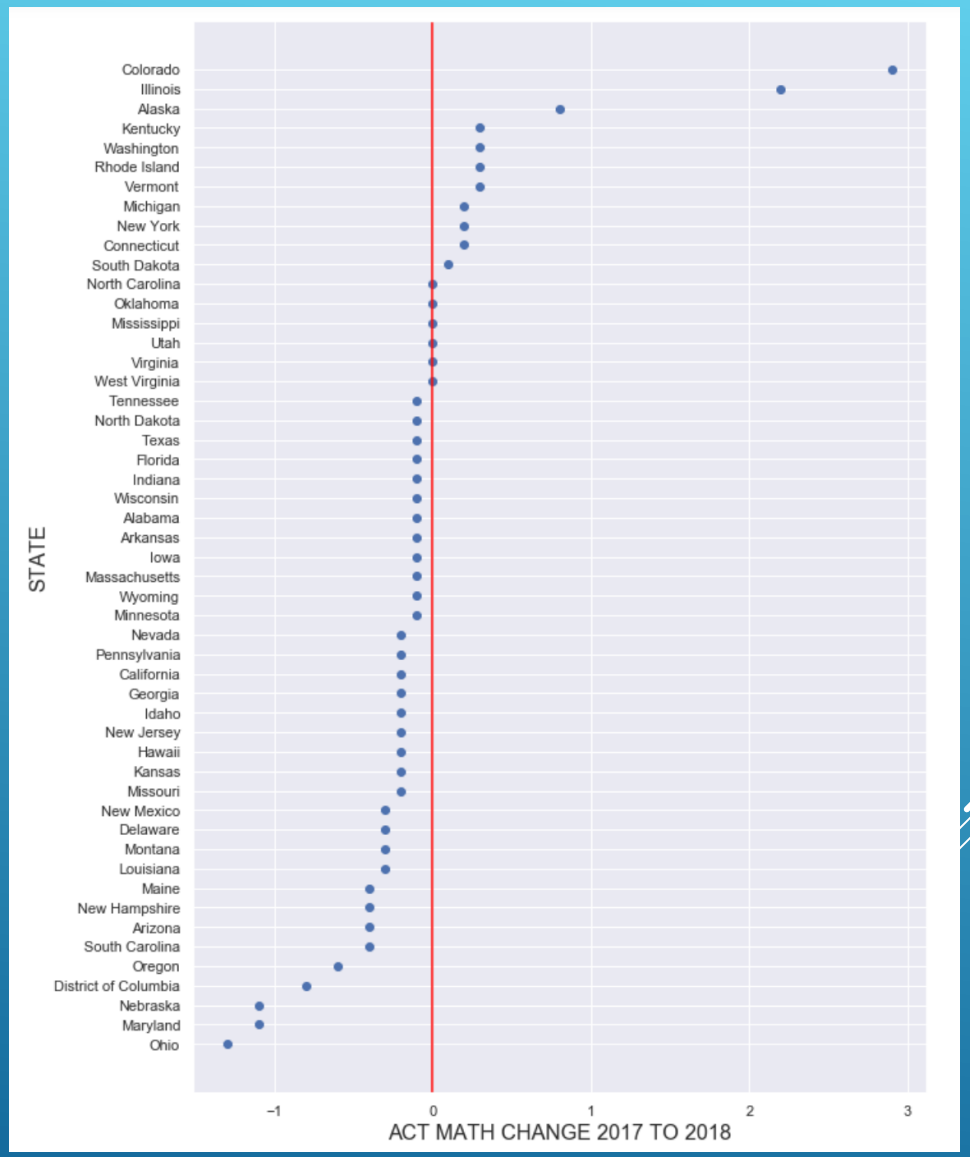
- From participation rates, it is not immediately obvious how many students took any test
- Participation rates alone do not fully explain changes in mean scores
- The number of students taking the test gives context to participation rates and changes in scores.

IMPACT OF CHANGE: PARTICIPATION

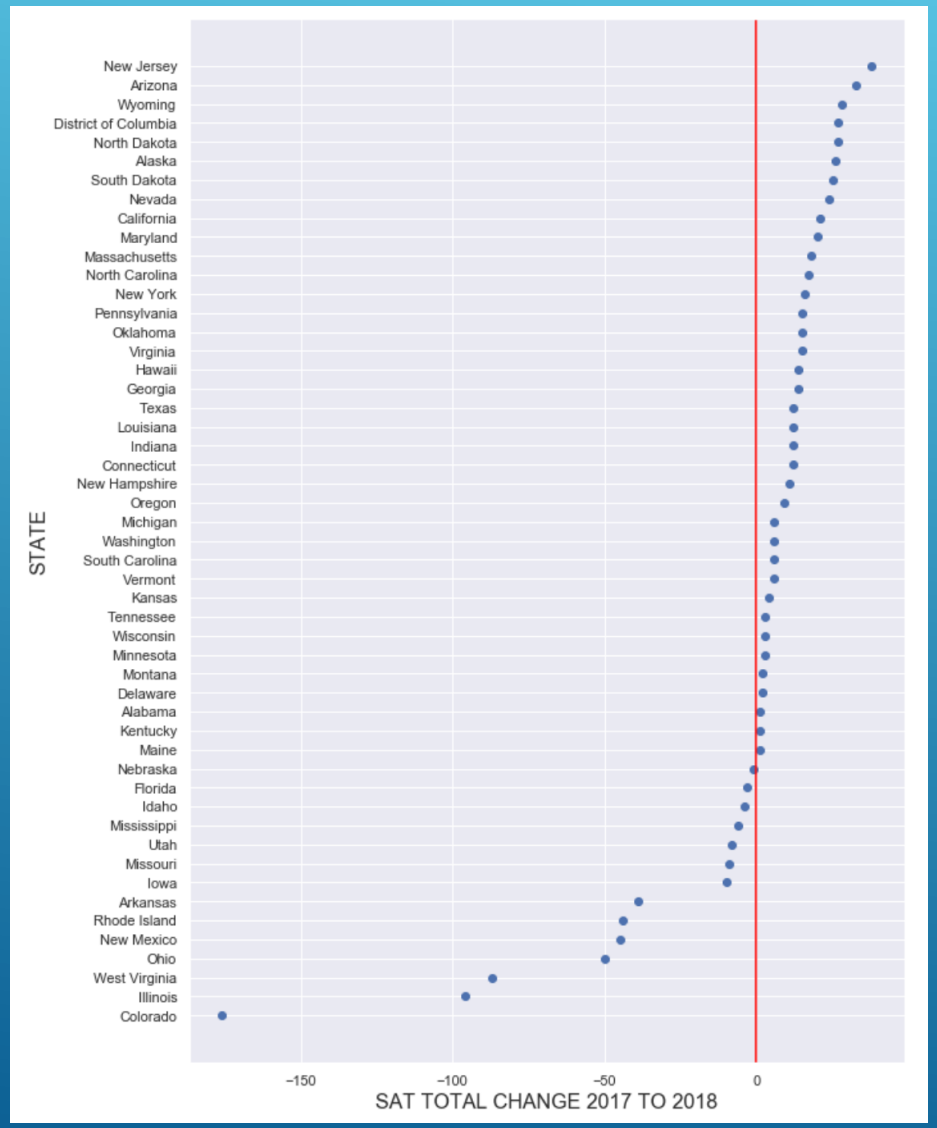


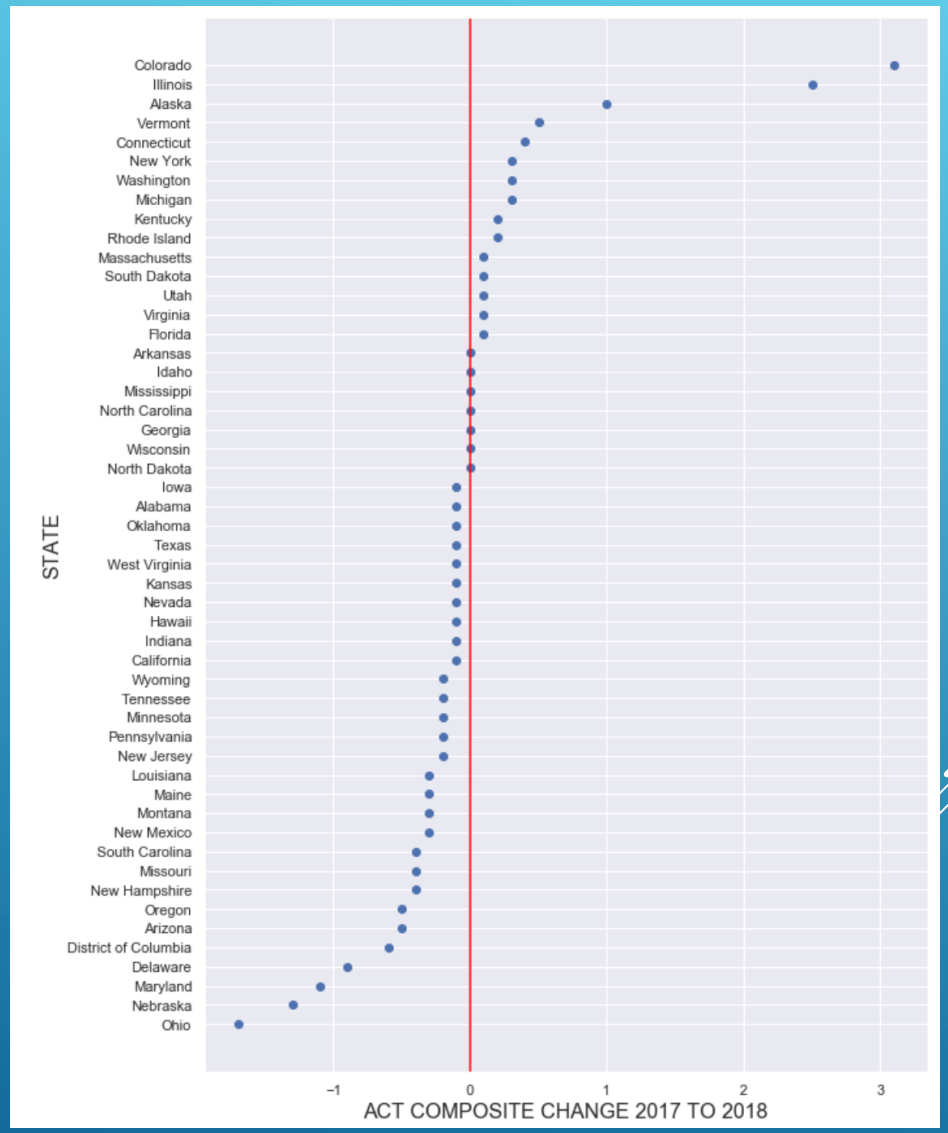


IMPACT OF CHANGE: MATHEMATICS



IMPACT OF CHANGE: TOTAL AND COMPOSITE





STATES TO WATCH

- Illinois
- Colorado
- Alaska

STATES TO WATCH

- Illinois
  - State government of Illinois made the SATs compulsory and started phasing out the ACT in 2016
  - Students were strongly incentivised:
    - State sponsorship
    - Tested at taker's convenience

STATES TO WATCH

- Colorado
  - Similar to Illinois, the state of Colorado signed a contract with the College Board to adopt and push the SAT in 2016.
  - Students were strongly incentivised:
    - State sponsorship
    - High convenience

STATES TO WATCH

- Alaska
  - Saw a large drop in ACT participation rates because the ACT was "losing ground" nationwide.

STATES TO WATCH

- Interesting observations:
  - All 3 states saw sharp dips in ACT participation
  - All 3 states saw the largest increases in mean scores for the ACTs

- This supports the idea that, to an extent:
  - Increased participation lowers overall means scores and
  - Decreased participation raises overall mean scores

CONCLUSION

- Changes to SAT test structure have had a major impact on the US education system.

- Whether it is positive and long-lasting is inconclusive for now.
  - More samples are required for accurate predictions.

RECOMMENDATIONS TO COLLEGE BOARD

- Continue reviewing how the new SAT format affects participation levels and test scores.
  - More data required to make more conclusive analysis
  - Factor in student populations (absolute numbers) in data to give context to participation levels
- To ensure future dominance, ensure higher adoption of SAT now
  - Obtain more government contracts for delivery of SAT exams

ONE STATE TO LOOK OUT FOR

- District of Columbia
  - DC's SAT participation rate saw a decrease with no perceptible corresponding increase in ACT participation rates.
- Possible that DC students may be taking alternative tests such as the ACT and PARCC.
- College Board could try negotiating with private colleges to offer more incentives to take SAT:
  - More scholarship money
  - Better perks
  - Prestige

KEY TAKEAWAYS

- Data loses its meaning without context.
  - Research can explain numbers, but numbers cannot tell a story by themselves.
- EDA helps you understand your data and what areas to research in order to make your analysis more accurate.
- The more similar the dependent variables of two datasets, the more equitable the comparison.
- Craft a good problem statement:
  - Essential for effective and efficient execution of EDA.

QUESTIONS?

THANK YOU!

REFERENCES

1. Business Insider: SAT is changing the format of its test in 2016
2. College Board's official report on the SAT re-design
3. https://www.act.org/content/dam/act/unsecured/documents/pdfs/R1670-college-readiness-benchmarks-2017-11.pdf
4. https://collegereadiness.collegeboard.org/about/scores/benchmarks
5. https://blog.prepscholar.com/average-sat-scores-by-state-most-recent
6. https://www.edweek.org/ew/section/multimedia/states-require-students-take-sat-or-act.html
7. https://www.chicagotribune.com/news/ct-illinois-chooses-sat-met-20160211-story.html
8. https://www.washingtonpost.com/education/2018/10/23/sat-reclaims-title-most-widely-used-college-admission-test/
9. https://www.edweek.org/ew/articles/2018/10/17/math-scores-slide-to-a-20-year-low.html
10. https://www.act.org/content/dam/act/secured/documents/cccr2018/National-CCCR-2018.pdf
11. https://www.edweek.org/ew/articles/2018/10/31/sat-scores-rise-as-number-of-test-takers.html