SAT & ACT Analysis

Problem Statement

- Data error analysis
  - To check through the entire dataset and compare with the official web
- Investigate trends in data
  - Finding out states with the highest and lowest participation rates for 2017/2018 SAT/ACT
  - Finding out states with the highest and lowest mean total and composite scores for 2017/2018 SAT/ACT
  - Finding out if there are states with 100% participation on a give test have a rate change year-to-year
  - Finding out if there are states with more than 50% participation on either year
  - Finding out if there are possible relations between scores and participation rate
- Recommendations
  - Suggest how College Board might increase participation for state with low participation rates

Error Analysis

- SAT Score minimum Math score is below the lower bound of 200
- ACT score for Science is abnormally low
- Converting Participation data type from String to float.
- Masking to filter out data types that are not float

```
df_sat_2017.describe()
```

	Evidence-Based Reading and Writing	Math	Total
count	51.000000	51.000000	51.000000
mean	569.117647	547.627451	1126.098039
std	45.666901	84.909119	92.494812
min	482.000000	52.000000	950.000000

```
df_act_2017.describe()
```

	English	Math	Reading	Science
count	52.000000	52.000000	52.000000	52.000000
mean	20.919231	21.173077	22.001923	21.040385
std	2.332132	1.963602	2.048672	3.151113
min	16.300000	18.000000	18.100000	2.300000

	State	Participation	Evidence-Based Reading and Writing	Math	Total
0	Alabama	6%	595	571	1166

```
# Find out which row cannot be converted to float
df_act_2017[~df_act_2017['Composite'].apply(isfloat)]
```

	State	Participation	English	Math	Reading	Science	Composite
51	Wyoming	100	19.4	19.8	20.8	20.6	20.2x

Finding out Min and Max

```
final.sort_values('sat_2017_participation')[['state', 'sat_2017_participation']].head(5)
```

	state	sat_2017_participation
34	North Dakota	2
24	Mississippi	2
15	Iowa	2
25	Missouri	3
44	Utah	3

```
final.sort_values('sat_2017_participation', ascending=False)[['state', 'sat_2017_participation']].head()
```

	state	sat_2017_participation
8	District of Columbia	100
22	Michigan	100
6	Connecticut	100
7	Delaware	100
29	New Hampshire	96

- State with lowest participation rates
  - 2017 ACT : **Maine** (8%)
  - 2018 ACT : **Maine** (7%)
  - 2017 SAT : **North Dakota** (2%)
  - 2018 SAT : **North Dakota** (2%)
- States with highest participation rates
  - 2017 ACT : **Colorado**, **Alabama**, **Kentucky**, **Wisconsin**, **Utah** ... (100%)
  - 2018 ACT : **Missouri**, **Alabama**, **Kentucky**, **Wisconsin**, **Utah** ... (100%)
  - 2017 SAT : **District of Columbia**, **Michigan**, **Connecticut**, **Delaware**, **New Hampshire** (100%)
  - 2018 SAT : **Colorado**, **Connecticut**, **Delaware**, **Michigan**, **Idaho** (100%)
- State with highest mean total/composite scores
  - 2017 ACT : **New Hampshire** (25.5)
  - 2018 ACT : **Connecticut** (25.6)
  - 2017 SAT : **District of Columbia** (950)
  - 2018 SAT : **District of Columbia** (977)
- State with lowest mean total/composite scores
  - 2017 ACT : **Nevada** (17.8)
  - 2018 ACT : **Nevada** (17.7)
  - 2017 SAT : **Minnesota** (1295)
  - 2018 SAT : **Minnesota** (1298)

States with high change in participation rate

```
# Filter states with 100% act participation
final[final['act_2017_participation'] > 90][['state','act_2017_participation', 'act_2018_participation']]
```

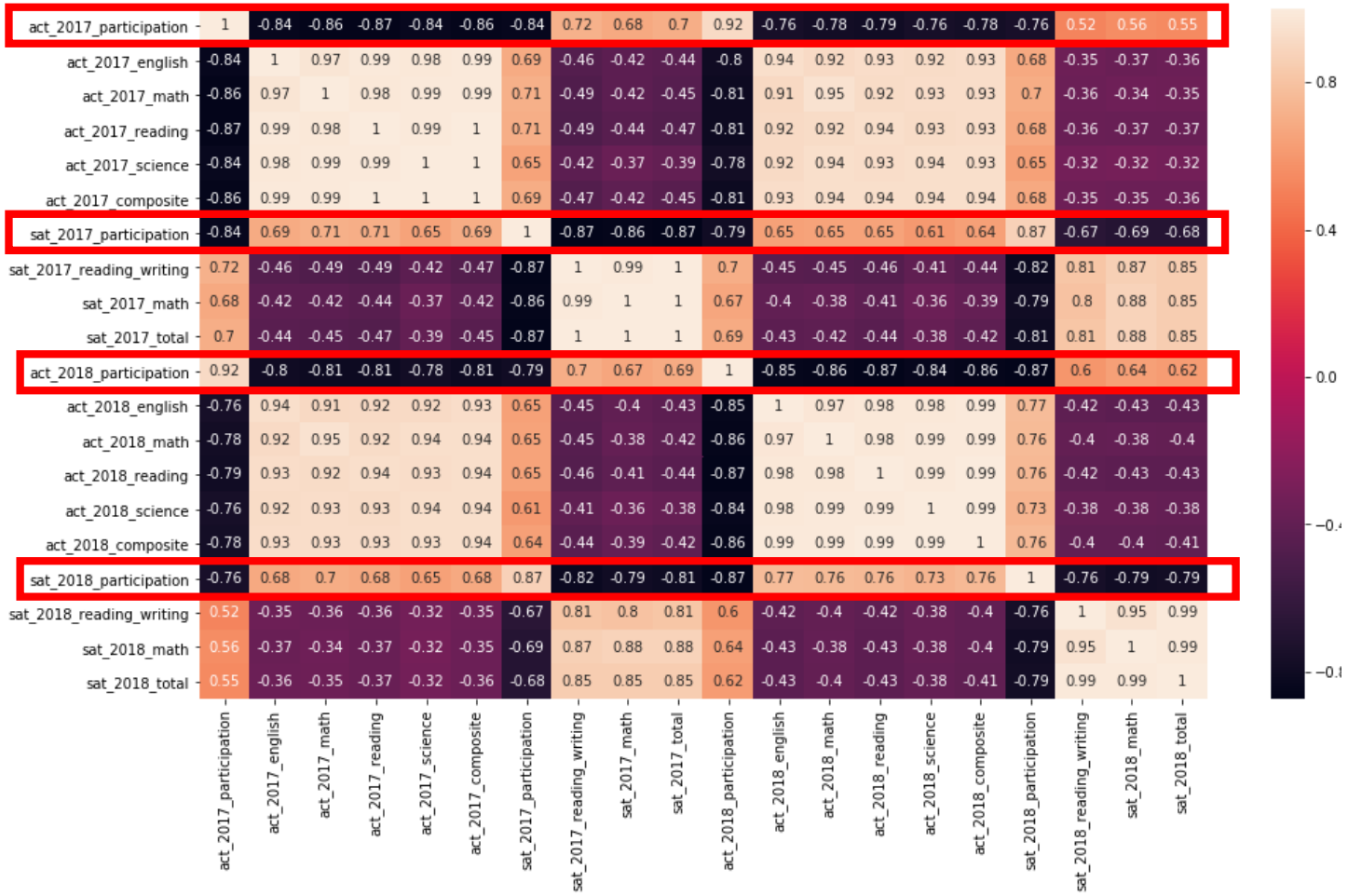
	state	act_2017_participation	act_2018_participation
0	Alabama	100	100
3	Arkansas	100	100
5	Colorado	100	30
13	Illinois	93	43

```
final[(final['sat_2017_participation'] < 50)
      & (final['sat_2018_participation'] > 90)][['state','sat_2017_participation', 'sat_2018_participation']]
```

	state	sat_2017_participation	sat_2018_participation
5	Colorado	11	100
13	Illinois	9	99

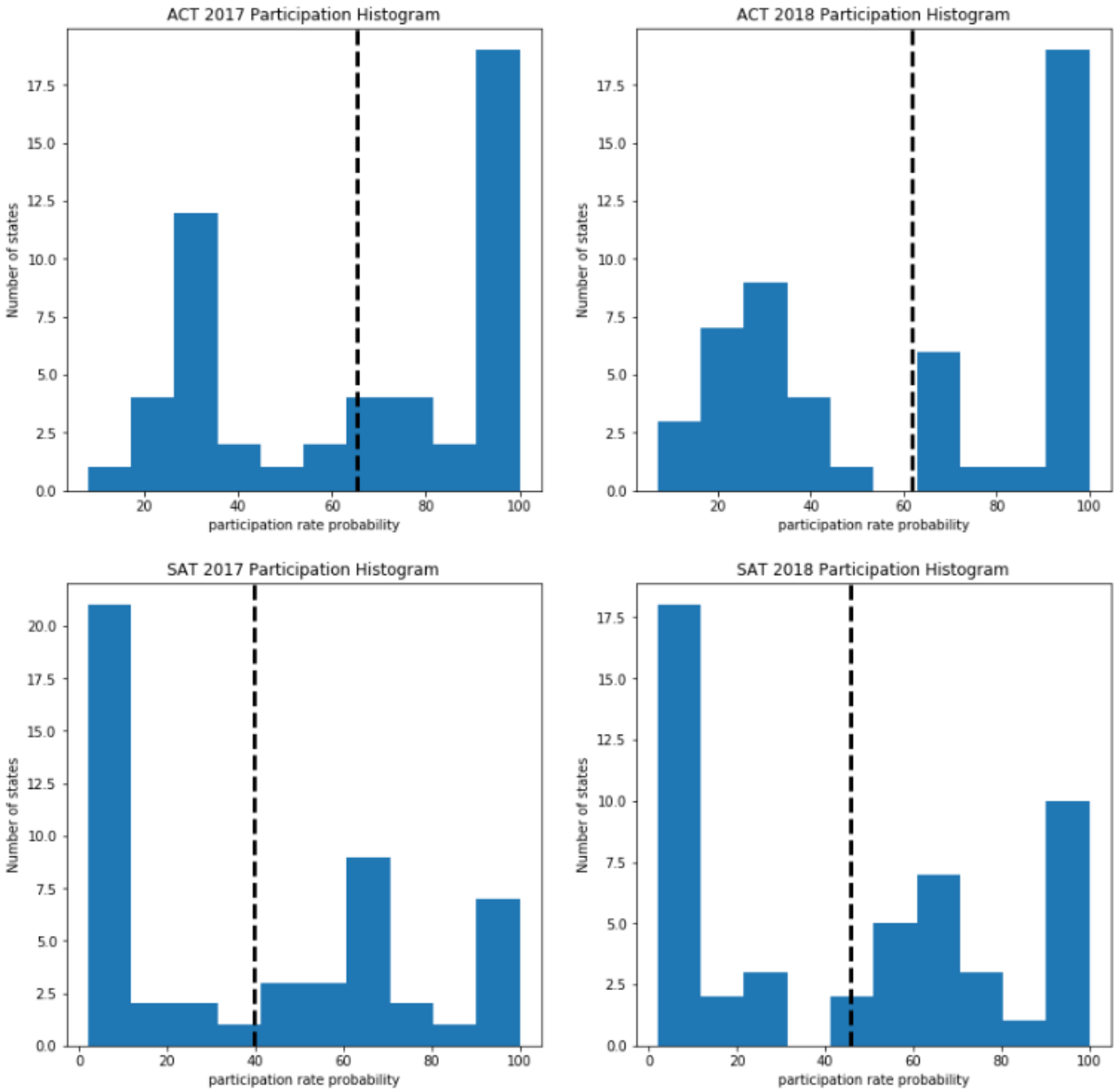
Correlation Heatmap

- Strong negative relationship between test participation and average score
- States with lower participation rate likely to see higher average scores than a state with higher participation rates on that same test.



Histograms

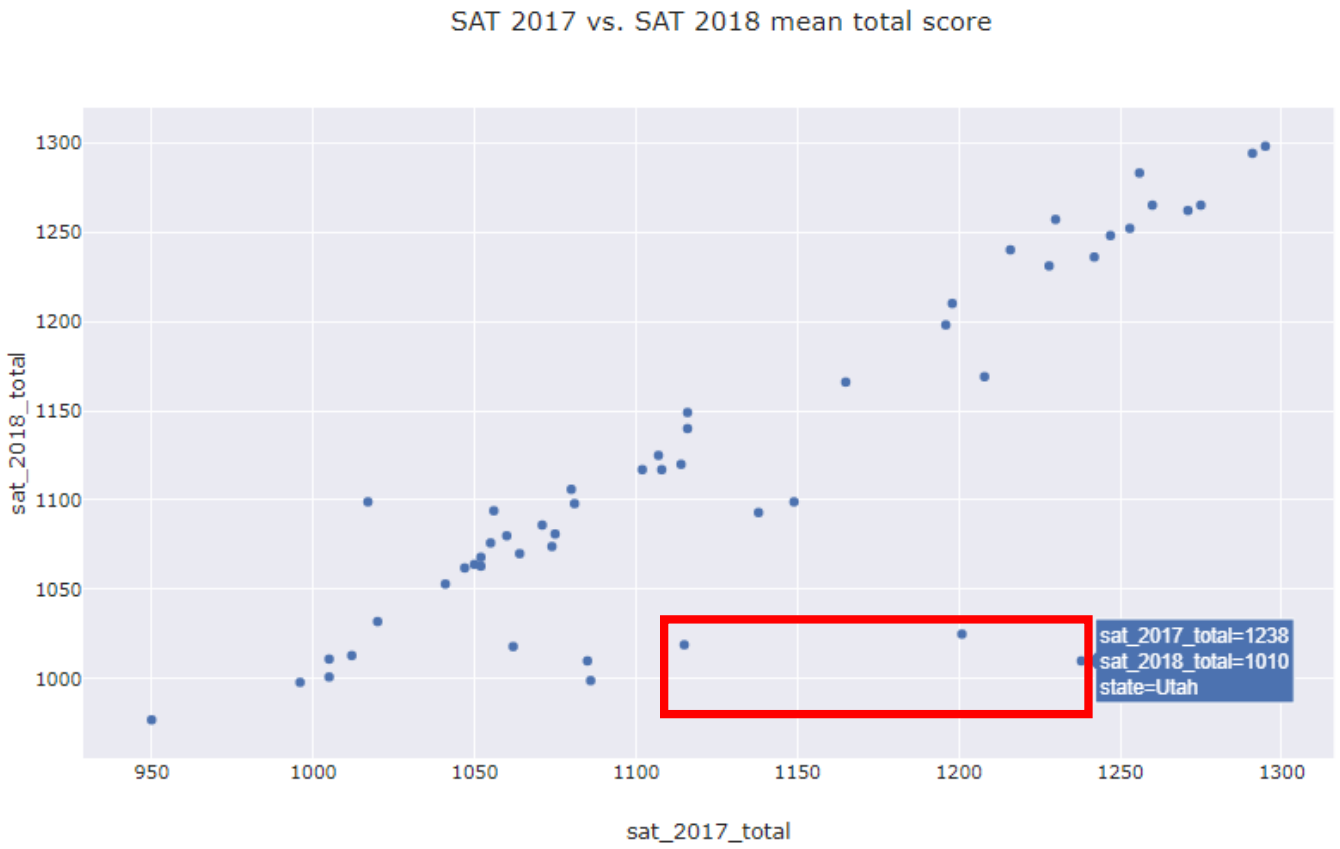
- Distribution for ACT participation rates are skewed left
- Distribution for SAT participation rates are skewed right for both years.
- ACT participation is higher than SAT.



Scatter Plots

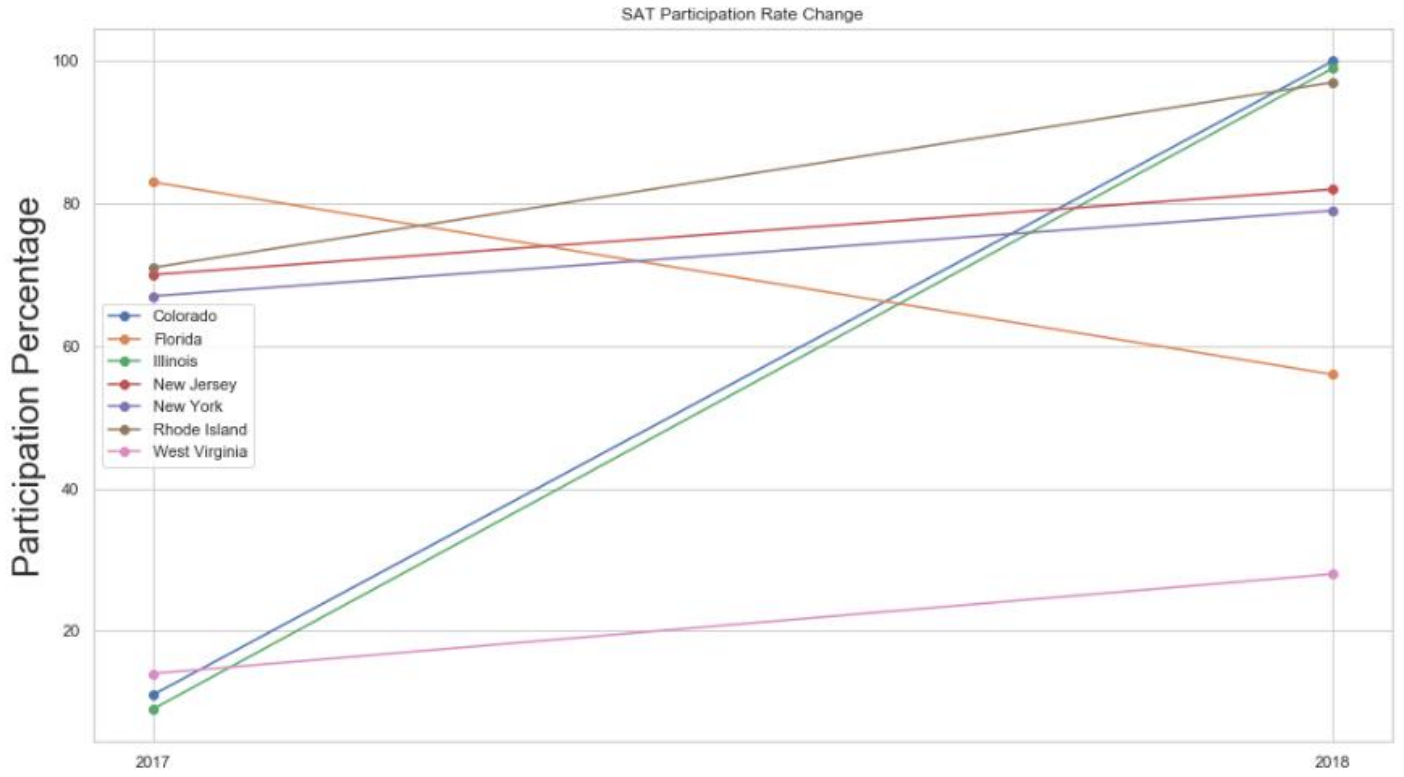
- 3 Outliers in the Scatter plots
  - Utah
  - Colorado
  - Illinois

- Huge decrease in 2017 SAT mean total score from 2018



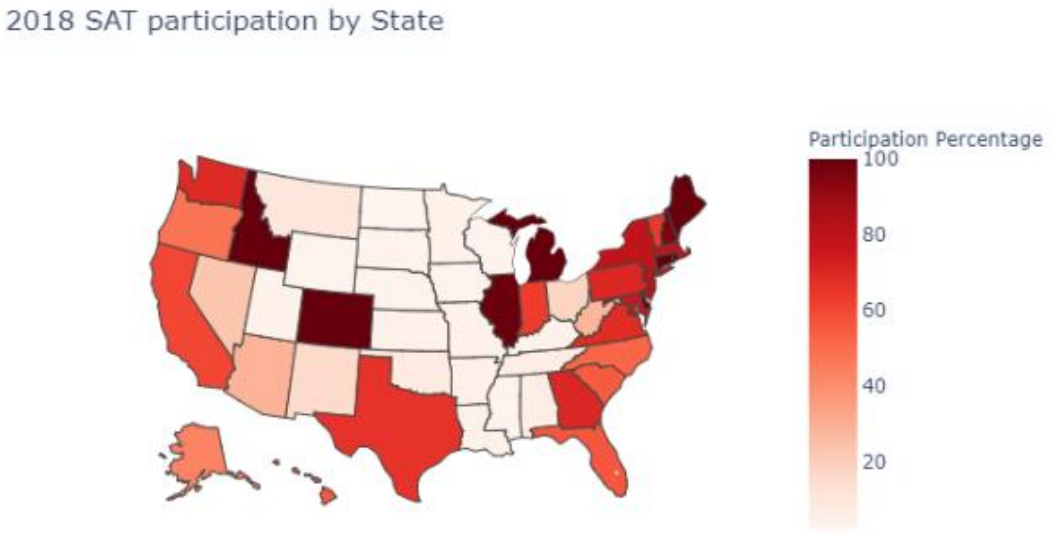
Slope Graph

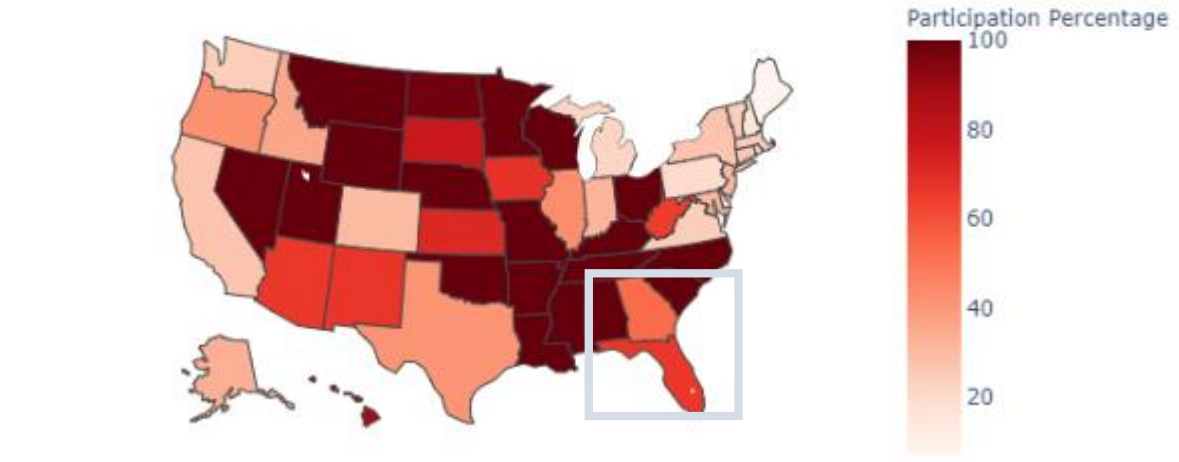
- Slope graph to find the rate of change in SAT and ACT 2017 to 2018 participation rates
- Colorado and Illinois has huge increase in SAT participation rates



Choropleth Map

- The colours of SAT and ACT are inverted
- Florida, Georgia and Hawaii have equal amount of participation rate for SAT and ACT





A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so we reject the null hypothesis. A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so we fail to reject the null hypothesis.

```
alpha = 0.05
```

```
# null hypothesis: x comes from a normal distribution
normal_distributed = {col: stats.normaltest(final[col])[1] for col in final.columns
                      if col!='state' and col!='state_code' and stats.normaltest(final[col])[1] > alpha }
normal_distributed
```

```
{'act_2017_english': 0.08434764489205682,
 'act_2017_science': 0.06388547443366324,
 'sat_2017_math': 0.056072998982864684,
 'sat_2018_math': 0.08866620599134843}
```

```
# null hypothesis: x comes from a normal distribution
not_normal_distributed = {col: stats.normaltest(final[col])[1] for col in final.columns
                          if col!='state' and col!='state_code' and stats.normaltest(final[col])[1] < alpha }
not_normal_distributed
```

```
{'act_2017_participation': 1.8018419611696254e-16,
 'act_2017_math': 0.03786805958606832,
 'act_2017_reading': 0.04396683079879635,
 'act_2017_composite': 0.03767739587231253,
 'sat_2017_participation': 4.028943345766873e-06,
 'sat_2017_reading_writing': 0.0009986795076584,
 'sat_2017_total': 0.008200230814088539,
 'act_2018_participation': 1.1798576998141957e-31,
 'act_2018_english': 0.011118664949764285,
 'act_2018_math': 0.008678423456739134,
 'act_2018_reading': 0.009055804952130748,
 'act_2018_science': 0.010625933240751445,
 'act_2018_composite': 0.012862760086159716,
 'sat_2018_participation': 3.739999415853251e-12,
 'sat_2018_reading_writing': 0.01822276093099164,
 'sat_2018_total': 0.014558724804979849}
```

Outside Research

- Colorado students' participation for ACT decreased 70% from 100% in 2017 to 30 % in 2018. These students went to take SAT and the SAT participation rate for Colorado increased 89% from 11% in 2017 to 100% in 2018. This is due to new contract acquired by SAT. This also result in decrease in the mean SAT total score, from 1201 to 1025 as well as an increase in the composite ACT score, from 20.8 to 23.9 from year 2017 to 2018.
- Similarly for Illinois students' participation for ACT decreased 53% from 93% in 2017 to 40 % in 2018. These students went to take SAT and the SAT participation rate increased 90% from 9% in 2017 to 99% in 2018. This is due to the new contract acquired by SAT, similar to Colorado. This also result in decrease in the mean SAT total score, from 1115to 1019 as well as an increase in the composite ACT score, from 21.4 to 23.9 from year 2017 to 2018.
- Florida, Georgia, Hawaii students have almost equal amount of students participating in SAT and ACT
- Rhode Island also has increase in SAT participation rates, from 71% to 97%
- Utah has a huge drop in SAT mean score, from 1238 to 1010

Conclusion

- Strong Competition between SAT and ACT test
- State department education have large impact on the participation rates
  - Impact by providing subsidies
  - Making it compulsory for all students in the state
- Collaboration with state education departments and colleges would lead to increase in SAT participation rates

- States with equal amount of students participating in SAT and ACT are recommended to look into for increase in SAT participation rates