

---

---

# Cleaning & Making Assumptions with SAT|ACT Data

A Study By Aakash Sharma

---



# SAT|ACT Analysis

## Problem:


- Combine SAT and ACT scores and participation rates from each state in the US.
- Identify trends to analyze likely factors influencing participation rates & scores in various states.

→ **Unexpected**  
Data Cleaning?

→ **Emotional**  
Stress Involved

→ **Simple**  
Yes, but tedious

# What is Data Cleaning



**Data cleaning** is detecting & correcting corrupt or inaccurate records, modifying, or deleting the dirty data so it can be used for more useful means

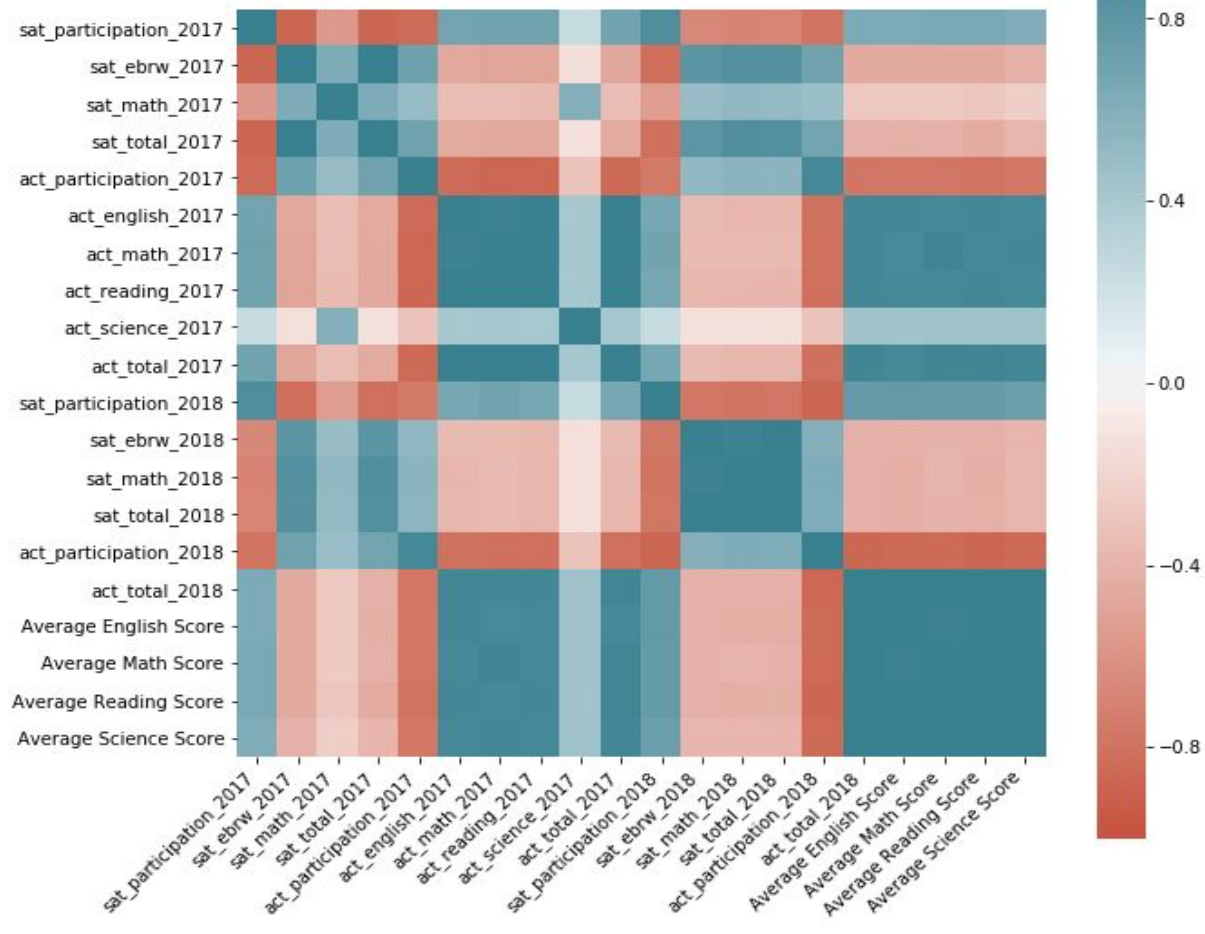


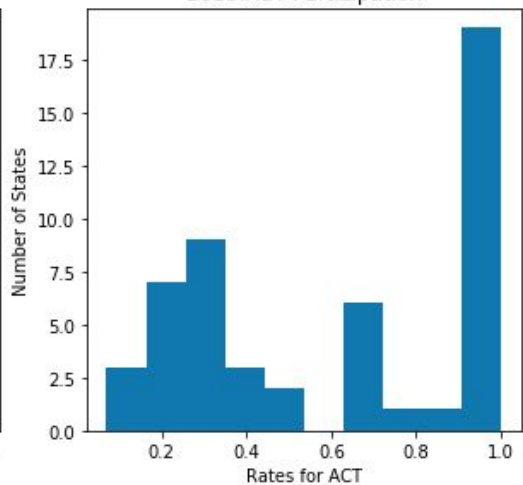
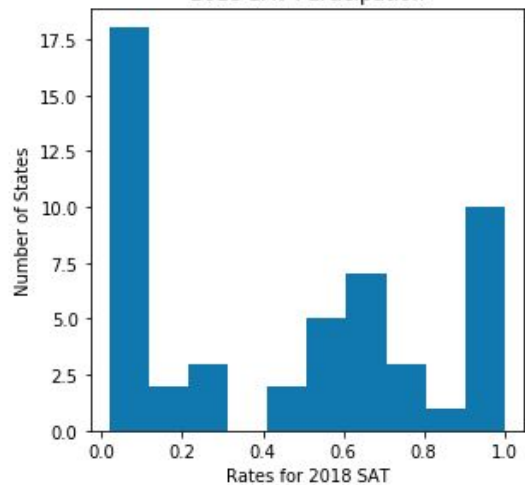
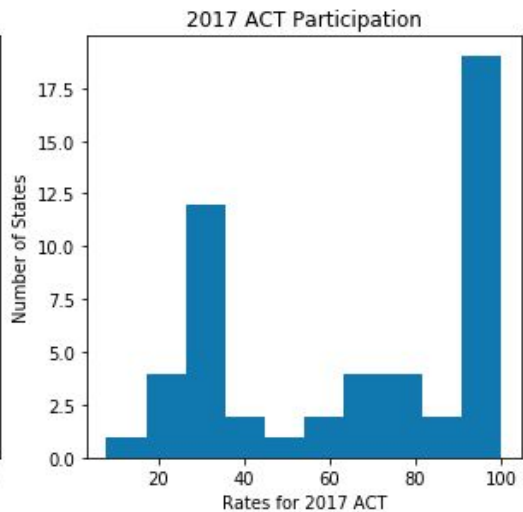
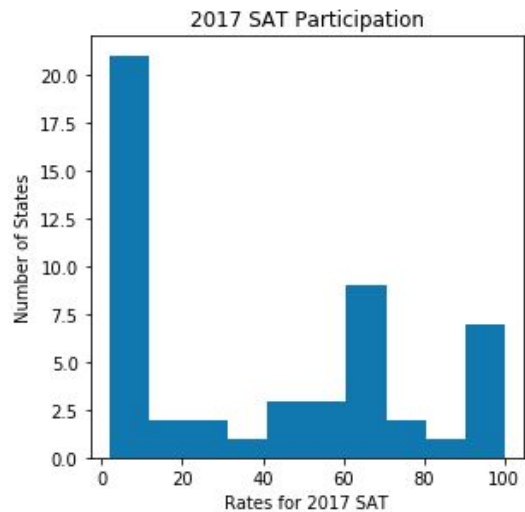
# Inconsistencies (With the Data)

- 'The National Column'
- Converting an object column to float
- Integer columns containing a string character
- Float columns had %

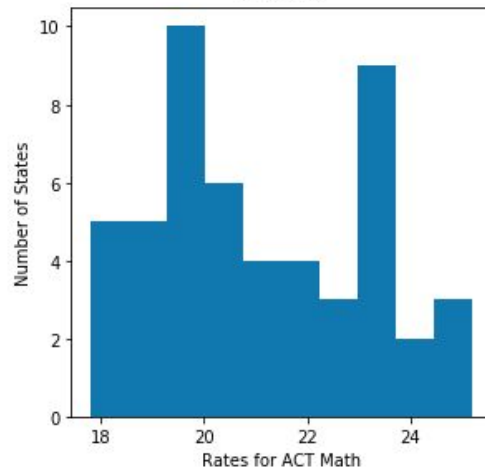
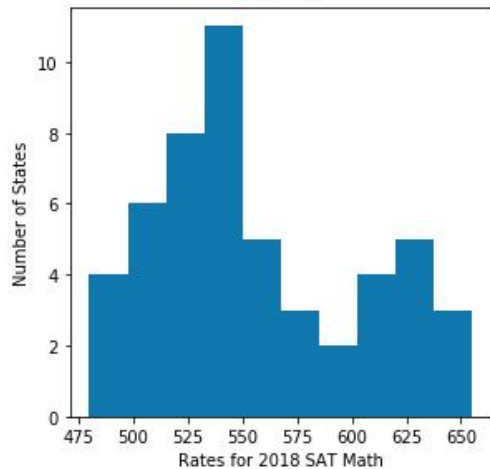
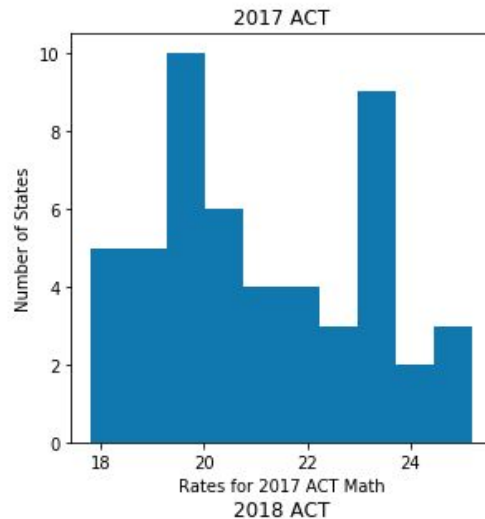
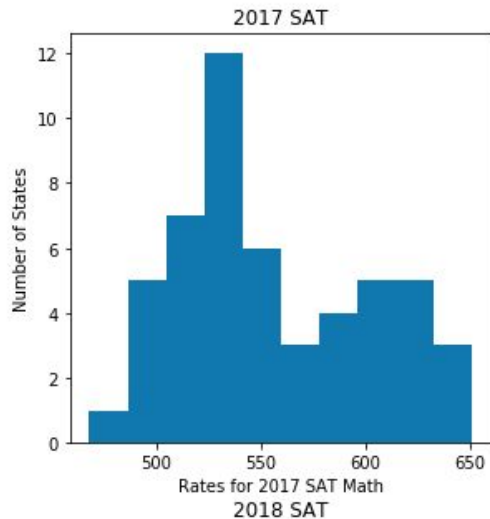
# Graphing the Data







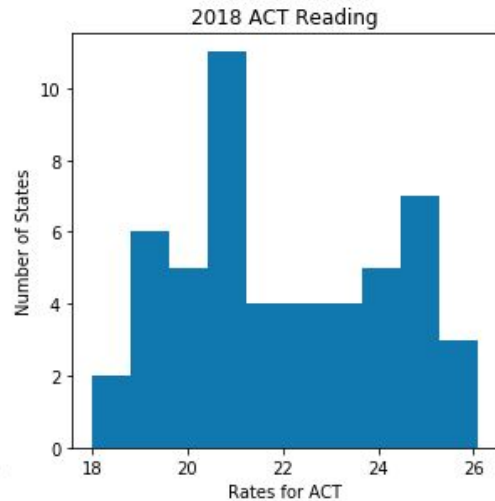
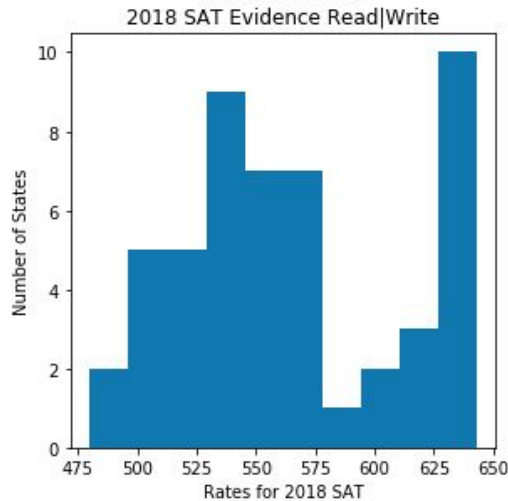
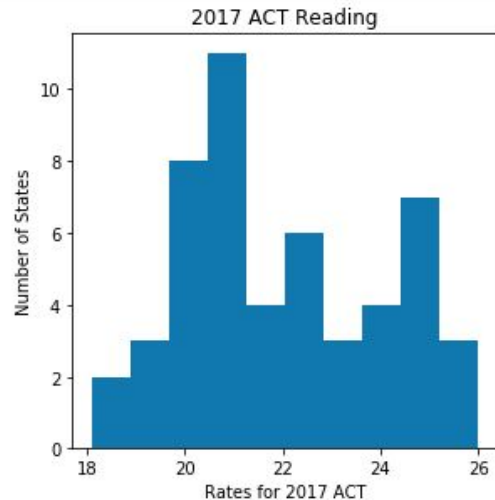
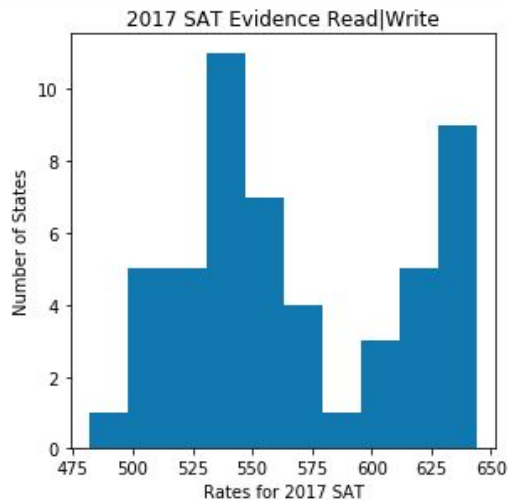
Participation rates  
for SAT & ACT



Math Scores

SAT & ACT



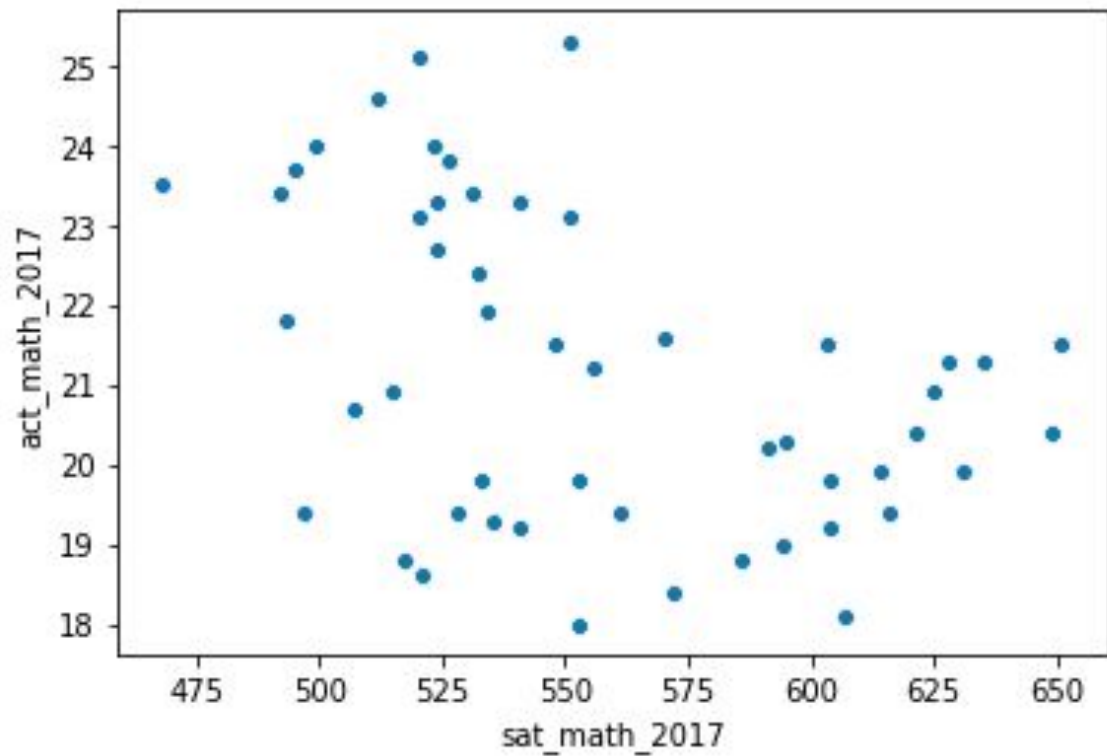


Reading/verbal

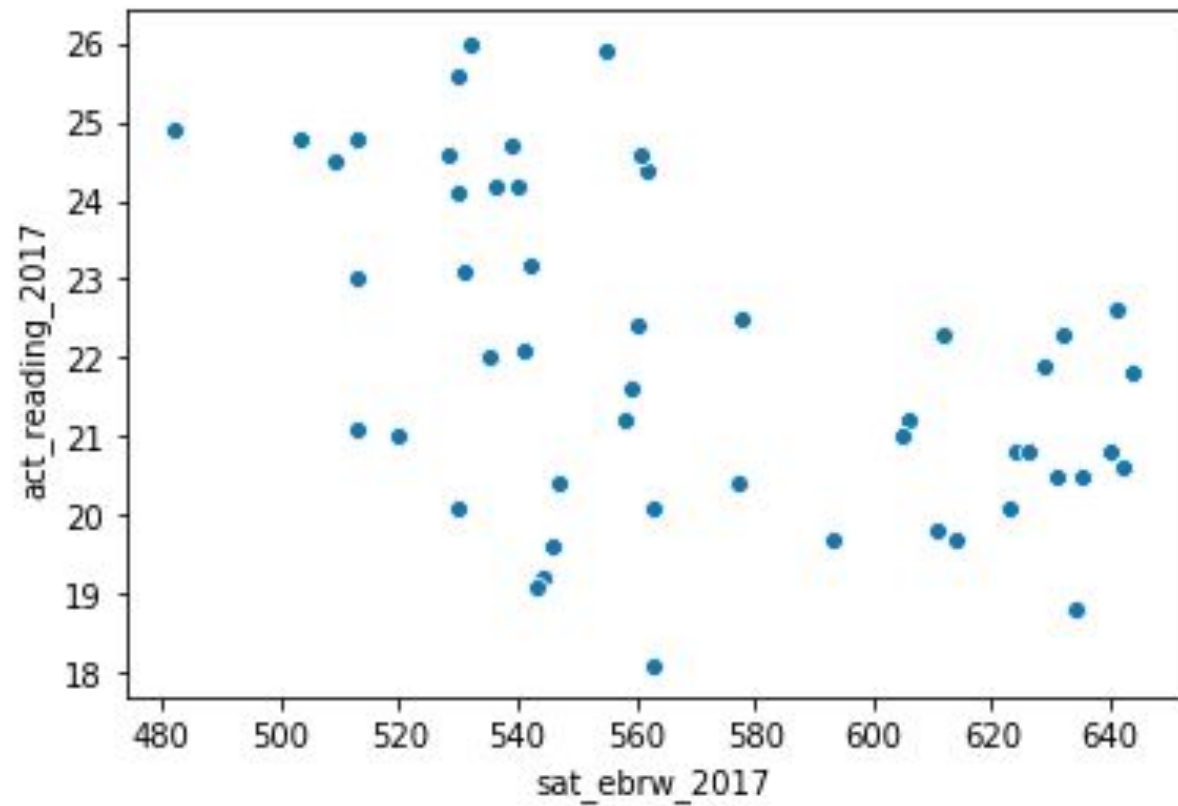
scores

SAT & ACT

- SAT vs. ACT math scores for 2017



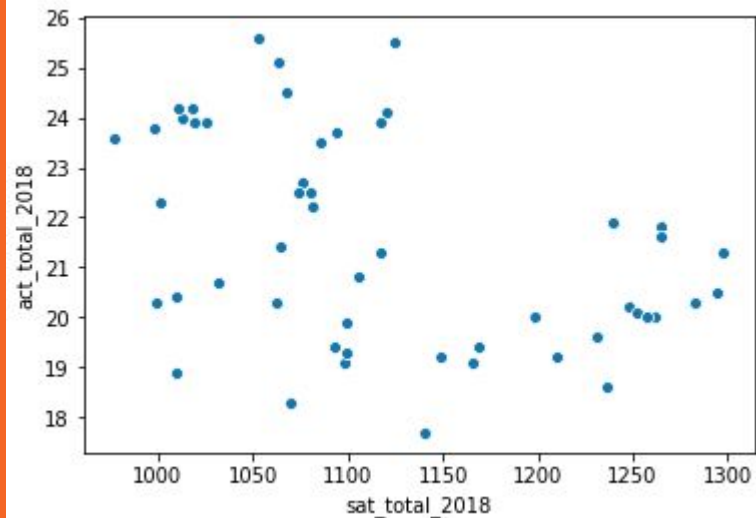
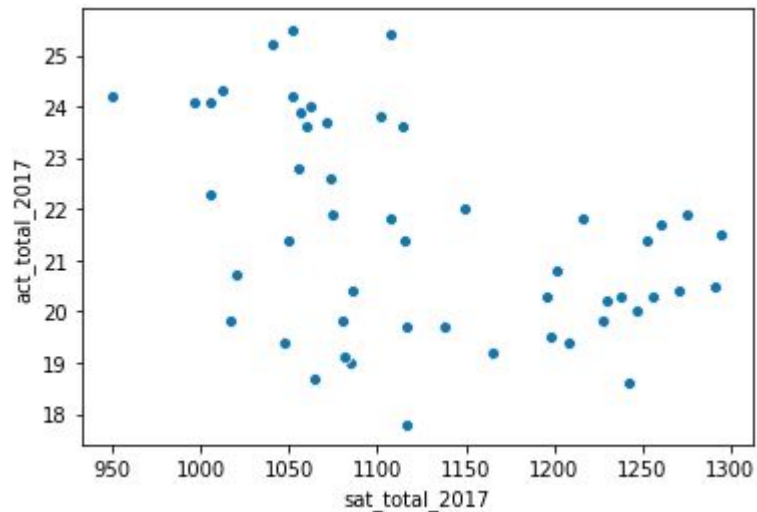
• SAT vs. ACT verbal/reading scores for 2017



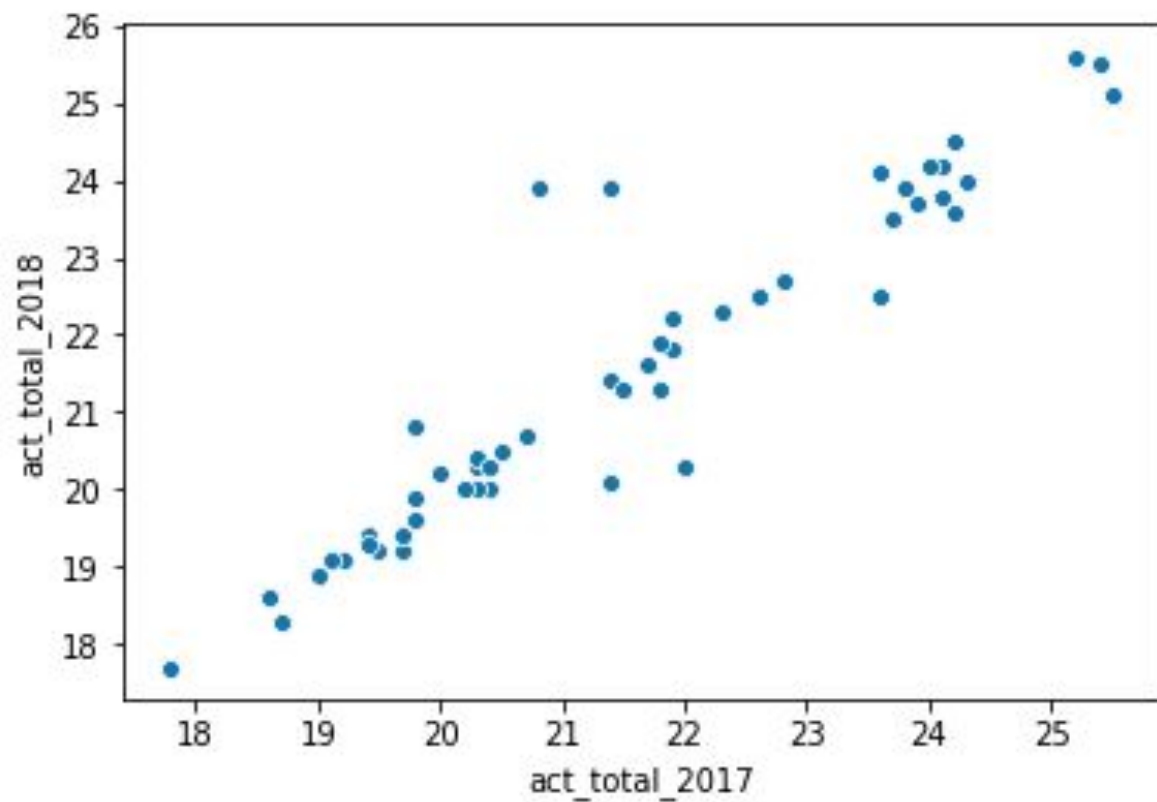
achieve  
more<sup>®</sup>

# SAT<sup>®</sup>

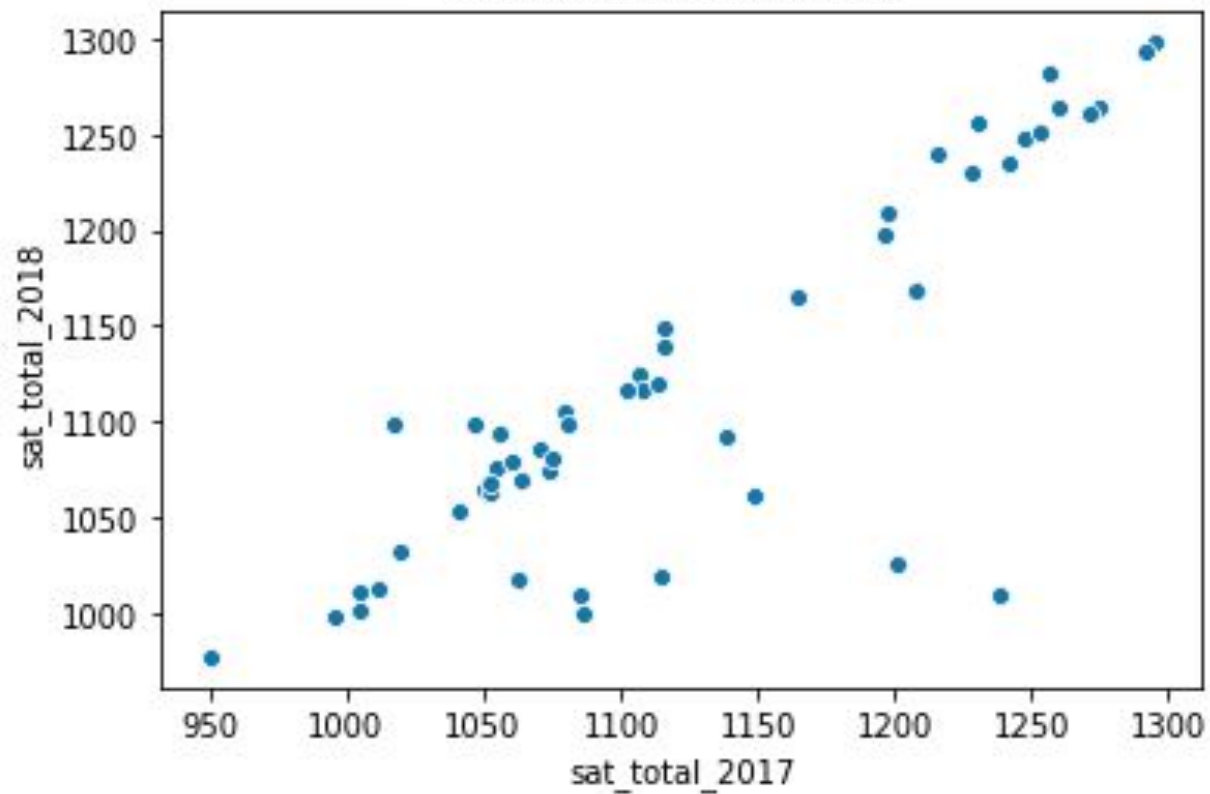
- **SAT vs. ACT total/composite scores for 2017**

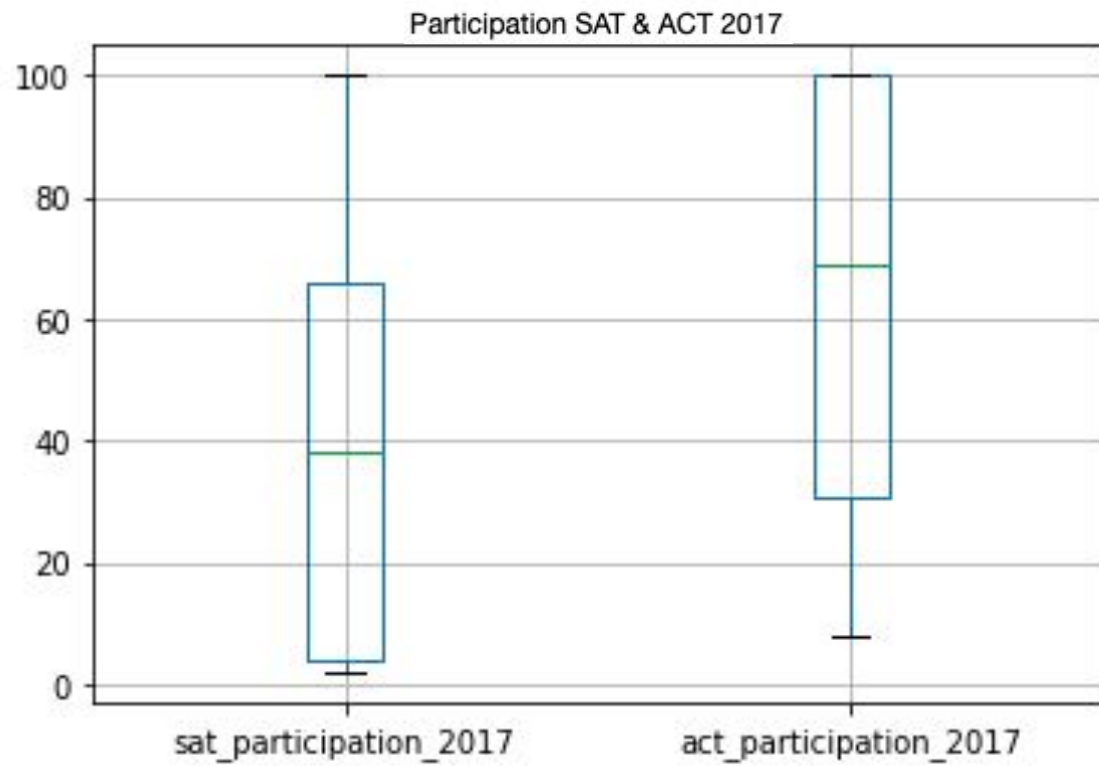


- **Composite scores for ACT 2017 vs. 2018**



- *Total scores for SAT 2017 vs. 2018*





Participation SAT & ACT 2018

