# Socioeconomic Factors vs. ACT Performance

Leiana Beatriz Omine

Seattle University

DATA 3320 – SP'23

# Introduction

- The Domain Problem: Is school performance predicted by socioeconomic factors?
- We are focusing on average student performance on the ACT or SAT exams that students take as part of the college application process and if socioeconomic factors may be used to predict school performance.

# Data Used

- EdGap includes information about average ACT or SAT scores for schools and several socioeconomic characteristics of the school district in 2016.
- National Center for Education Statistics includes basic information about each school.
- Education Commission of the States provides reports that compare educational policies across 50 states from 2016-2017.
- Education Week provides a table with the 2016-2017 data of the states that require SAT/ACT and other exams.

# Questions

What is the relationship between the requirement for students to take the ACT/SAT and the average scores on these tests across different states?

Since the data does not include all the states, does this affect the relationship between ACT performance and socioeconomic factors? Can the inclusion of ACT/SAT requirements help counteract this bias?

# Analysis Methods

- Multiple Linear Regression
- Best Subset Selection

# Multiple Linear Regression

- High r-squared value indicates good fit.
- Low p-value (p<0.5) indicates statistically significant.
- Low standard error indicates important predictor.

---

- R-squared with 63% variance for `average_act`.
- Lowest significance in the model are `median_income` and `percent_married`.
- Lowest standard of error is `percent_lunch` with 0.108.

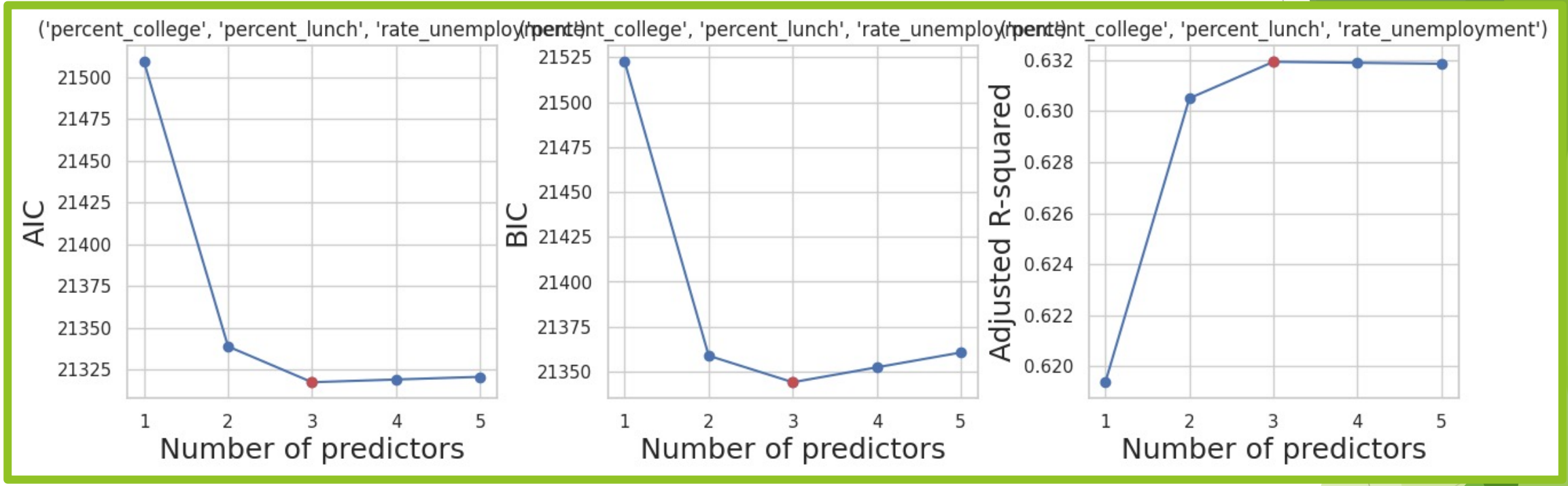
OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | average_act | R-squared: | 0.632 |
| Model: | OLS | Adj. R-squared: | 0.632 |
| Method: | Least Squares | F-statistic: | 1985. |
| Date: | Wed, 07 Jun 2023 | Prob (F-statistic): | 0.00 |
| Time: | 01:07:14 | Log-Likelihood: | -10654. |
| No. Observations: | 5781 | AIC: | 2.132e+04 |
| Df Residuals: | 5775 | BIC: | 2.136e+04 |
| Df Model: | 5 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 22.7774 | 0.154 | 147.937 | 0.000 | 22.476 | 23.079 |
| percent_lunch | -7.7132 | 0.108 | -71.109 | 0.000 | -7.926 | -7.501 |
| median_income | 1.067e-06 | 1.34e-06 | 0.799 | 0.425 | -1.55e-06 | 3.69e-06 |
| percent_married | -0.0961 | 0.150 | -0.640 | 0.522 | -0.390 | 0.198 |
| percent_college | 1.5641 | 0.177 | 8.842 | 0.000 | 1.217 | 1.911 |
| rate_unemployment | -2.0735 | 0.453 | -4.575 | 0.000 | -2.962 | -1.185 |

| | | | |
|---|---|---|---|
| Omnibus: | 738.501 | Durbin-Watson: | 2.014 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 2546.433 |
| Skew: | 0.632 | Prob(JB): | 0.00 |
| Kurtosis: | 5.996 | Cond. No. | 1.36e+06 |

# Best Subset Selection



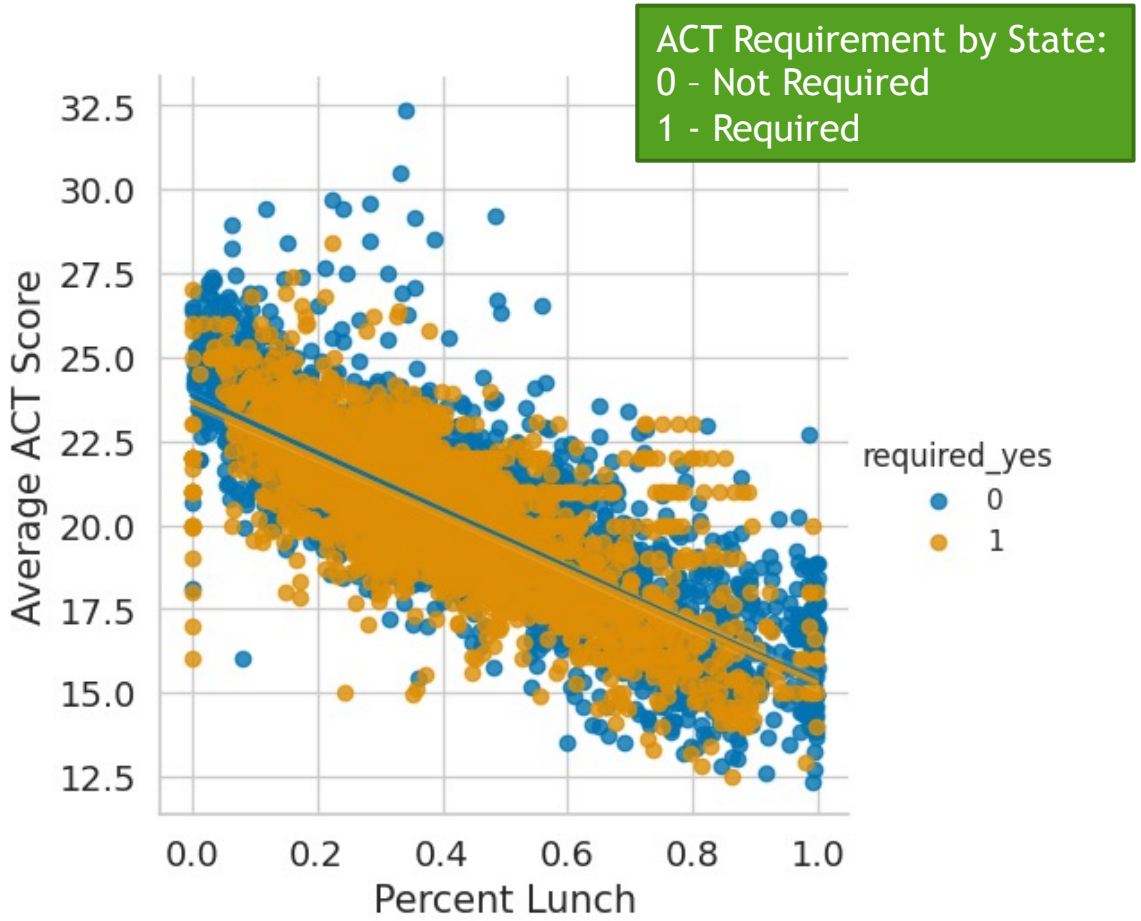

- Higher r-squared values indicate better performance.
- Lower AIC/BIC values indicate better fit.
- The best predictors are colored red.
  - The most effective predictor is `percent_lunch`.
  - Best subset is `percent_college`, `percent_lunch`, and `rate_unemployment` on `average_act`.

# Additional Step Analysis

- What is the relationship between the requirement for students to take the ACT/SAT and the average scores on these tests across different states?

---

- There is a strong negative relationship between average ACT and percent lunch.
- Students who receive free or reduced lunch are more likely to be associated with low average ACT.

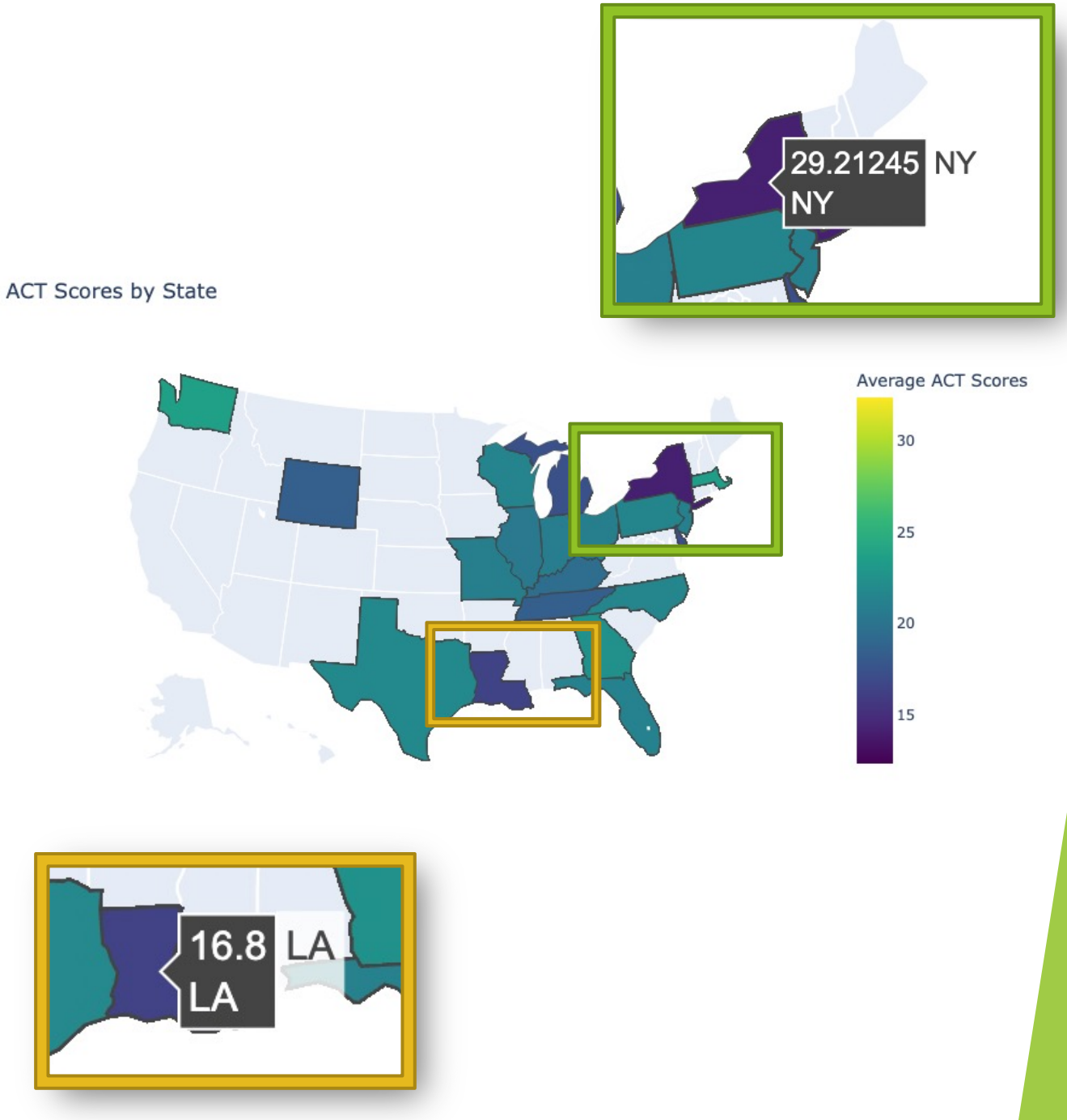


# Additional Step Analysis

- What is the relationship between the requirement for students to take the ACT/SAT and the average scores on these tests across different states?

---

- New York had the highest average ACT score at 29.2.
- Louisiana had the lowest average ACT score at 16.8.



# Additional Step Analysis

- Since the data does not include all the states, does this affect the relationship between ACT performance and socioeconomic factors? Can the inclusion of ACT/SAT requirements help counteract this bias?

Required by State

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | average_act | R-squared: | 0.620 |
| Model: | OLS | Adj. R-squared: | 0.620 |
| Method: | Least Squares | F-statistic: | 4714. |
| Date: | Wed, 07 Jun 2023 | Prob (F-statistic): | 0.00 |
| Time: | 01:07:20 | Log-Likelihood: | -10748. |
| No. Observations: | 5781 | AIC: | 2.150e+04 |
| Df Residuals: | 5778 | BIC: | 2.152e+04 |
| Df Model: | 2 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 23.8396 | 0.045 | 529.607 | 0.000 | 23.751 | 23.928 |
| required_yes == 1[T.True] | -0.1228 | 0.042 | -2.945 | 0.003 | -0.204 | -0.041 |
| percent_lunch | -8.4940 | 0.087 | -97.094 | 0.000 | -8.665 | -8.322 |

| | | | |
|---|---|---|---|
| Omnibus: | 710.866 | Durbin-Watson: | 1.744 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 2272.882 |
| Skew: | 0.630 | Prob(JB): | 0.00 |
| Kurtosis: | 5.801 | Cond. No. | 5.47 |

Not Required by State

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | average_act | R-squared: | 0.620 |
| Model: | OLS | Adj. R-squared: | 0.620 |
| Method: | Least Squares | F-statistic: | 4714. |
| Date: | Wed, 07 Jun 2023 | Prob (F-statistic): | 0.00 |
| Time: | 01:07:20 | Log-Likelihood: | -10748. |
| No. Observations: | 5781 | AIC: | 2.150e+04 |
| Df Residuals: | 5778 | BIC: | 2.152e+04 |
| Df Model: | 2 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 23.7168 | 0.047 | 499.916 | 0.000 | 23.624 | 23.810 |
| required_yes == 0[T.True] | 0.1228 | 0.042 | 2.945 | 0.003 | 0.041 | 0.204 |
| percent_lunch | -8.4940 | 0.087 | -97.094 | 0.000 | -8.665 | -8.322 |

| | | | |
|---|---|---|---|
| Omnibus: | 710.866 | Durbin-Watson: | 1.744 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 2272.882 |
| Skew: | 0.630 | Prob(JB): | 0.00 |
| Kurtosis: | 5.801 | Cond. No. | 5.87 |

# Additional Step Conclusion

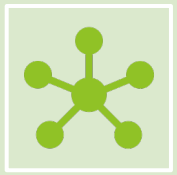
We found a strong negative relationship between `average_act` and `percent_lunch`.

We created regression models using these predictors to determine if the ACT requirement is useful for our analysis. However, we found that there is only a 0.001 difference between the two groups where ACT was required and not required.

We can conclude that the ACT requirement has little to no impact when finding relationships between ACT performance and socioeconomic factors.

# Conclusion

There are limitations to our analysis due to missing significant predictors.



The data provided did not include all the states and does not take the recent SAT changes from 2018 into account. These factors may have affected our results.

Using the states included in this data set as a predictor is not useful when finding relationships between ACT performance and socioeconomic factors.