

Final Project

MinJae Jo

2025-12-02

Introduction

Introduction

For this project, I am looking at the question: “Does the size or competitiveness of a college affect its degree attainment rate?” I chose C150_4 as my response variable because it shows the share of students who finish their degree within 150% of the usual time. It is one of the clearer numbers in the dataset that relates to graduation.

For the explanatory variable, I am using SAT_AVG. The SAT score is not exactly the same as college size, but schools with higher SAT averages usually have more selective admissions and sometimes bigger or more competitive environments. So I thought it could work as a way to compare colleges that are different in scale or difficulty.

Both variables are continuous, so I am planning to use a linear model to see if they move together in some way. I think this topic is interesting. People often talk about whether large schools or hard-to-select schools help students succeed, but it is hard to know without seeing the data in person. I would like to see what the CollegeScorecard dataset actually shows.

Preprocessing

```
## grad_rate_150      sat_avg
## Min.      :0.0000   Min.      : 564
## 1st Qu.   :0.3229   1st Qu.   :1044
## Median    :0.4944   Median    :1116
## Mean      :0.4881   Mean      :1131
## 3rd Qu.   :0.6453   3rd Qu.   :1195
## Max.      :1.0000   Max.      :1558
## NA's      :4703     NA's      :5743
```

Visualization

Summary Statistics

Data Analysis

Conclusion