A Glimpse on Federalizing Public Education

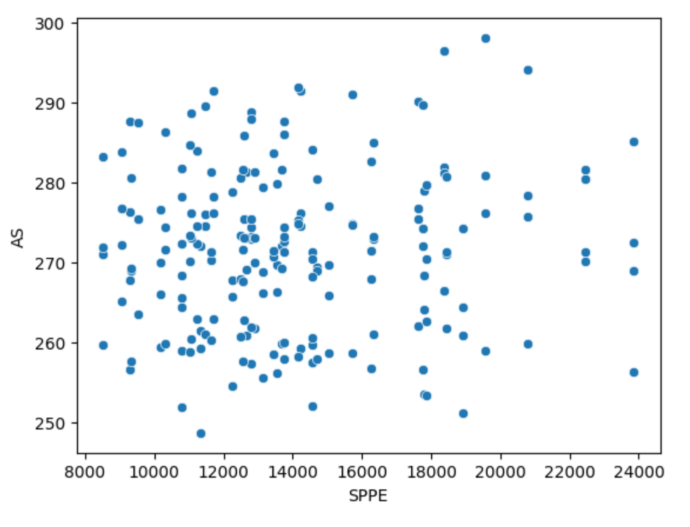
This is meant as a hypothetical test, based solely upon theory and existing data. The political ramifications of such are not discussed, as it provides more of a glimpse into a potential alternative theoretical pathway to addressing education rather than a realistic possibility of being implemented. Various Python packages are used, and you will find integrated Tableau visualizations as well as an exported SQL database created to host the code.

Created by Kolbe Alexandre Dumas with assistance of ChatGPT 4

The data is sourced from the NAEP, BEA, and NCES, and contains the following: data on 8th grade achievement scores by state (denoted by variable AS), broken into demographics of Ethnicity, Subject (math and reading), and Reduced Lunch status. It also contains data on nationwide Per-Pupil Expenditure (PPE) by state, and state Regional Price Parities (RPP) data to be used to standardize PPE into a value that can be compared accross different costs of living - this variable is SPPE. The demographic data has been aggregated into four categories: Not White or Asian and Reduced Lunch, Not White or Asian and No Reduced Lunch, White or Asian and Reduced Lunch, and White or Asian and No Reduced Lunch.

```
/Users/kolbedumas/miniconda3/lib/python3.10/site-packages/openpyxl/worksheet/header_footer.py:48: UserWarning: Cannot parse header or footer so it will be ignored
  warn("""Cannot parse header or footer so it will be ignored""")
```

The following is an initial scatter plot of SPPE vs AS, as the graph shows, there does not seem to be any correlation between a state's spending on education and achievement scores.

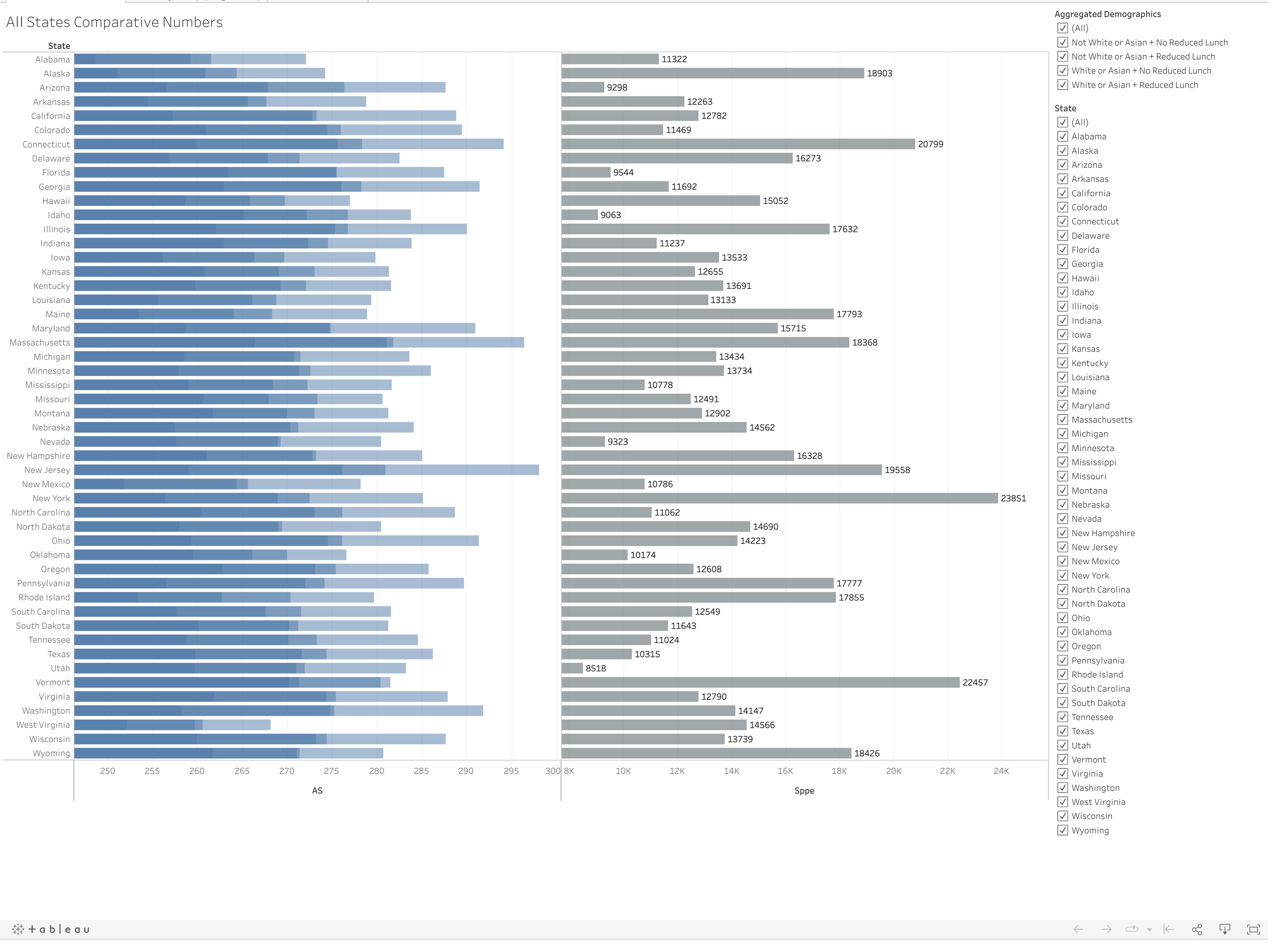


To take it a step further, we can see from the following regression analysis that SPPE is not statistically significant;however, we see that the demographic groups are. Not White or Asian + Reduced lunch groups see a significant drop in test scores compared to White or Asian + No Reduced Lunch; a difference in roughly 25 points.

```
Summary for Model 1
                            OLS Regression Results
==============================================================================
Dep. Variable:                     AS   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.776
Method:                 Least Squares   F-statistic:                     173.6
Date:                Mon, 04 Sep 2023   Prob (F-statistic):           4.20e-63
Time:                        12:51:42   Log-Likelihood:                -594.35
No. Observations:                 200   AIC:                             1199.
Df Residuals:                     195   BIC:                             1215.
Df Model:                           4
Covariance Type:            nonrobust
==================================================================================================================
                                                          coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                                 215.4854      1.112    193.763      0.000     213.292     217.679
SPPE                                                    0.0002   9.65e-05      1.566      0.119   -3.92e-05       0.000
Aggregated_Demographics_Not White or Asian + Reduced Lunch     41.2910      0.649     63.659      0.000      40.012      42.570
Aggregated_Demographics_White or Asian + Reduced Lunch         53.4441      0.649     82.396      0.000      52.165      54.723
Aggregated_Demographics_Not White or Asian + No Reduced Lunch  54.2986      0.649     83.713      0.000      53.019      55.578
Aggregated_Demographics_White or Asian + No Reduced Lunch      66.4517      0.649    102.449      0.000      65.172      67.731
==============================================================================
Omnibus:                        5.560   Durbin-Watson:                   1.038
Prob(Omnibus):                  0.062   Jarque-Bera (JB):                5.442
Skew:                          -0.305   Prob(JB):                       0.0658
Kurtosis:                       3.530   Cond. No.                     7.90e+19
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 6.66e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Below are three visualizations crafted in Tableau that can be filtered through to see specific data on state AS scores for the various demographics as well as the state's SPPE for reference.

```
/var/folders/ym/r8jvn1xx27v64dct1pcf_vs00000gn/T/ipykernel_4750/2601254607.py:2: DeprecationWarning: Importing display from IPython.core.display is deprecated since IPython 7.14, please import from IPython display
  from IPython.core.display import display, HTML
```



To get an idea of the disparity differences between the highest and lowest performing groups in different states, a new column "disparity" was created to sort by. We can see below the states who are closest in terms of how well thier groups do and those who are the furthest.

Out[94]:

Aggregated_Demographics	State	Not White or Asian + Reduced Lunch	White or Asian + No Reduced Lunch	Disparity
44	Vermont	270.218895	281.560002	11.341107
47	West Virginia	252.119631	268.215036	16.095405
35	Oklahoma	259.480271	276.607048	17.126777
10	Hawaii	258.639853	277.001610	18.361757
11	Idaho	265.169212	283.819191	18.649979
49	Wyoming	261.702866	280.764622	19.061756
25	Montana	261.749282	281.261542	19.512260
24	Missouri	260.717270	280.631348	19.914078
15	Kansas	260.846375	281.362538	20.516163
13	Indiana	263.008043	283.930549	20.922505

Out[95]:

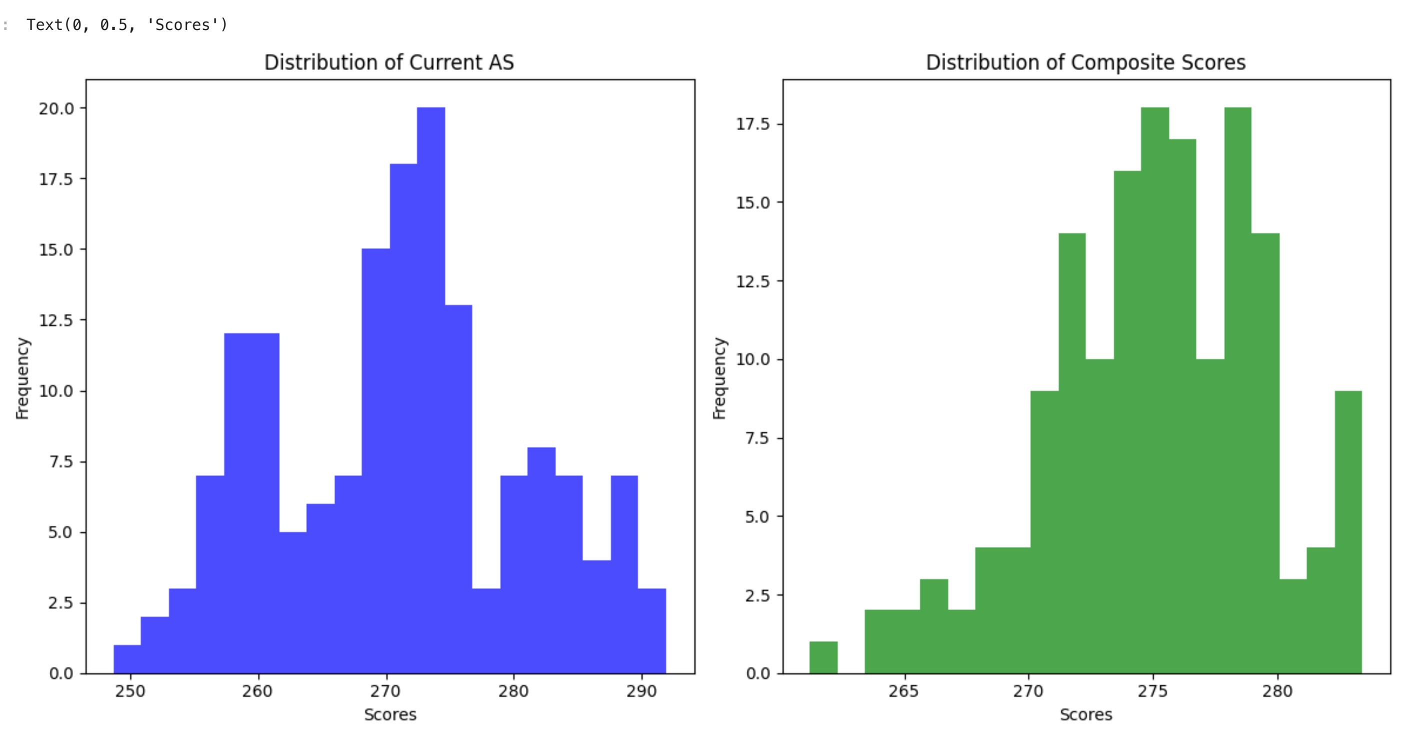
Aggregated_Demographics	State	Not White or Asian + Reduced Lunch	White or Asian + No Reduced Lunch	Disparity
31	New York	256.393260	285.144028	28.750768
20	Massachusetts	266.474012	296.474094	30.000082
2	Arizona	256.557819	287.659526	31.101707
4	California	257.309570	288.859906	31.550336
34	Ohio	259.319183	291.401303	32.082120
19	Maryland	258.616407	291.002860	32.386453
37	Pennsylvania	256.564104	289.718137	33.154034
46	Washington	258.228494	291.892710	33.664216
6	Connecticut	259.900224	294.174783	34.274559
29	New Jersey	259.047161	298.082595	39.035434

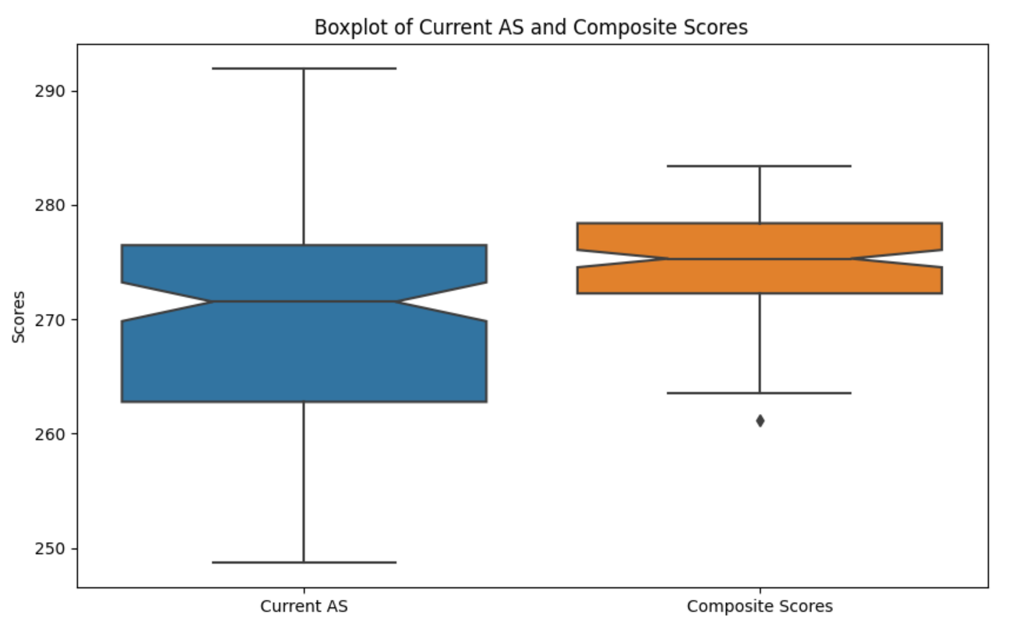
From the data we have, we can see there are essentially a few key factors; some states are very good in terms of overall scoring, however, they lack when it comes to narrowing the achievement gap as the disparity between the top performing groups and the bottom performing groups is large. There are also states who do fairly well at minimizing the achievement gap, but their overall scores lack. The idea is the following; if education were to be federalized, and thorough research was done on the top 5 states in each of these categories (lowest disparity + highest average AS score), could you combine the approaches to education these states take into one singular approach that combines them? This model attempts to do so, giving more weight to reducing the achievement score (0.7 to 0.3).

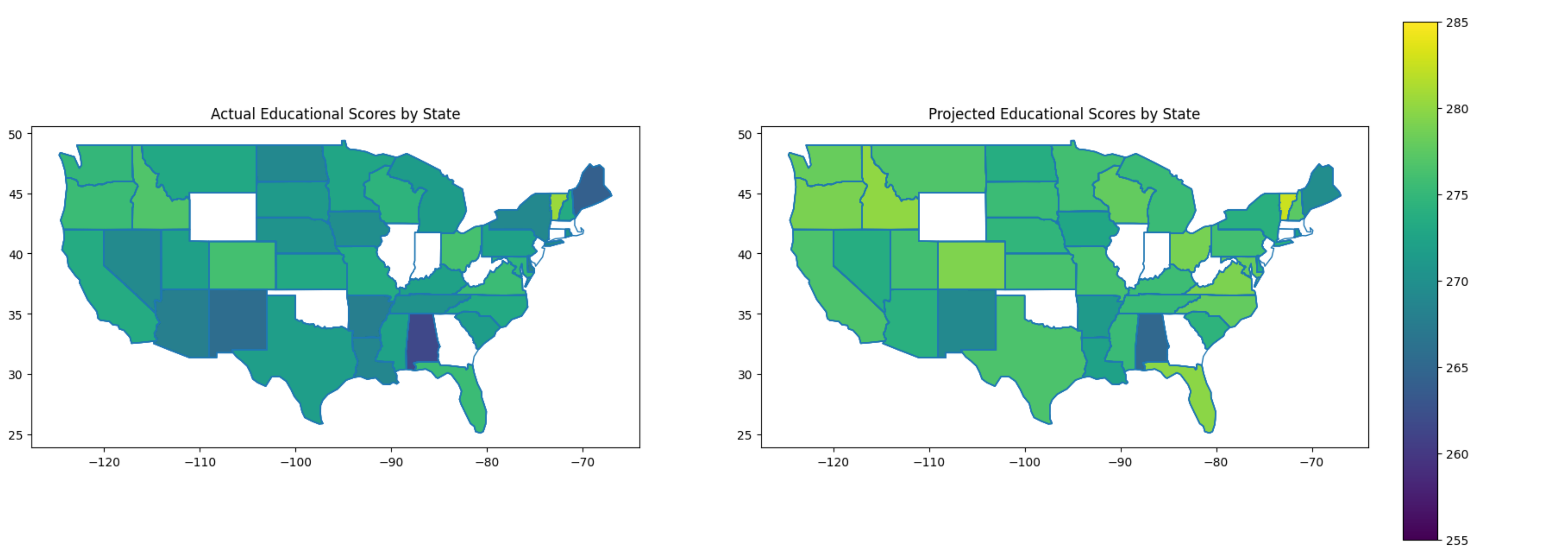
```
               AS  Projected_AS  Reduced_Disparity_AS
count  160.000000    160.000000            160.000000
mean   271.057189    278.074690            268.229545
std      9.806241      3.558134              9.825686
min    248.705659    267.420858            246.666044
25%    262.780043    276.542093            261.059433
50%    271.529317    278.337541            268.653634
75%    276.473223    280.983220            273.934683
max    291.892710    282.906949            285.845104

Current Scores Descriptive Statistics:
count    160.000000
mean     271.057189
std        9.806241
min      248.705659
25%      262.780043
50%      271.529317
75%      276.473223
max      291.892710
Name: AS, dtype: float64
Composite Scores Descriptive Statistics:
count    160.000000
mean     275.121146
std        4.438226
min      261.194414
25%      272.261057
50%      275.303230
75%      278.395199
max      283.417952
Name: Composite_Score, dtype: float64
Summary for the Filtered Model
                            OLS Regression Results
==============================================================================
Dep. Variable:                     AS   R-squared:                       0.830
Model:                            OLS   Adj. R-squared:                  0.826
Method:                 Least Squares   F-statistic:                     189.9
Date:                Mon, 04 Sep 2023   Prob (F-statistic):           1.19e-58
Time:                        12:51:42   Log-Likelihood:                -449.82
No. Observations:                 160   AIC:                             909.6
Df Residuals:                     155   BIC:                             925.0
Df Model:                           4
Covariance Type:            nonrobust
==================================================================================================================
                                                          coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                                 217.7134      1.077    202.094      0.000     215.585     219.841
SPPE                                                -8.018e-05   9.66e-05     -0.830      0.408      -0.000       0.000
Aggregated_Demographics_Not White or Asian + Reduced Lunch     41.8448      0.621     67.348      0.000      40.617      43.072
Aggregated_Demographics_White or Asian + Reduced Lunch         53.9417      0.621     86.817      0.000      52.714      55.169
Aggregated_Demographics_Not White or Asian + No Reduced Lunch  54.9149      0.621     88.384      0.000      53.688      56.142
Aggregated_Demographics_White or Asian + No Reduced Lunch      67.0119      0.621    107.853      0.000      65.785      68.239
==============================================================================
Omnibus:                        4.948   Durbin-Watson:                   1.035
Prob(Omnibus):                  0.084   Jarque-Bera (JB):                4.812
Skew:                          -0.294   Prob(JB):                       0.0902
Kurtosis:                       3.612   Cond. No.                     7.24e+19
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 5.93e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Summary for the Projected Model
                            OLS Regression Results
==============================================================================
Dep. Variable:        Composite_Score   R-squared:                       0.369
Model:                            OLS   Adj. R-squared:                  0.353
Method:                 Least Squares   F-statistic:                     22.66
Date:                Mon, 04 Sep 2023   Prob (F-statistic):           9.46e-15
Time:                        12:51:42   Log-Likelihood:                -428.14
No. Observations:                 160   AIC:                             866.3
Df Residuals:                     155   BIC:                             881.7
Df Model:                           4
Covariance Type:            nonrobust
==================================================================================================================
                                                          coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                                 221.0459      0.941    234.966      0.000     219.188     222.904
SPPE                                                 -8.77e-05   8.44e-05     -1.039      0.300      -0.000     7.9e-05
Aggregated_Demographics_Not White or Asian + Reduced Lunch     51.4864      0.543     94.892      0.000      50.415      52.558
Aggregated_Demographics_White or Asian + Reduced Lunch         55.1155      0.543    101.580      0.000      54.044      56.187
Aggregated_Demographics_Not White or Asian + No Reduced Lunch  55.4075      0.543    102.118      0.000      54.336      56.479
Aggregated_Demographics_White or Asian + No Reduced Lunch      59.0365      0.543    108.807      0.000      57.965      60.108
==============================================================================
Omnibus:                       19.331   Durbin-Watson:                   0.412
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               23.158
Skew:                          -0.792   Prob(JB):                     9.36e-06
Kurtosis:                       3.981   Cond. No.                     7.24e+19
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 5.93e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

The model shows what the numbers could look like in this theoretical sitatution; we see average scores raise, and the disparity lowers significantly compared to what it was before. While top scores might come down a bit, this would be a necessary price to pay. Below are visualizations of the current numbers and the projected numbers. A random forest was also ran to get a potential SPPE value that would be standard.

Out[98]: `Text(0, 0.5, 'Scores')`








```
Predictions:
0    11302.051174
1    11392.482799
2    12004.317988
3    11322.036970
4    11302.051174
Name: Aggregate_SPPE, dtype: float64
```