Factors Impacting College Enrollment Rate of Georgia High School Graduates

Brandon Lee, Carolina Segovia,
Daniel Chen & Margaret Wharton

August 3, 2021

Questions We Wanted to Answer:

What variables impact college enrollment of Georgia High School Graduates?

- Does a school district's revenue & expenditure impact college enrollment?
- Does academic performance impact college enrollment?
- Does economic advantage impact college enrollment?
- Do SAT & ACT scores impact college enrollment?

What are the best ways for districts to improve college enrollment of graduates?

Why We Chose Our Topic:

- We know that a college degree has become increasingly important
- **90.6%** of Georgia students are graduating from high school
- **67.8%** of graduates are enrolling in college upon graduation

Data Sources:

Governor's Office of Student Achievement Downloadable Data

(all reports used were 2018 - 2019 school year files)

- Post-Secondary C11 Report
- Ga Milestones EOC Assessment (by grade)
- Revenues and Expenditures
- Dropout Rate 7-12
- Graduation Rate (4-Year Cohort)
- ACT Scores (Recent)
- SAT Scores (Recent)

*https://gosa.georgia.gov/dashboards-data-report-card/downloadable-data

Data Exploration & Cleanup:

- Main database used to compare data to was the Post-Secondary C11 Report

  (Students graduating as well as number of students enrolled in a postsecondary institution within 16 months of graduation.)

- Comparison data used throughout analysis was the School Code, Percent of High School Graduates that entered into a Postsecondary Institute
- Remove data that's not an integer (TFS/NaN)
- Sort values based on independent variable
- Merge everything to dependent variable

cleaning > ind_var > ind_var_clean.py > ind_var_clean

```
1  def ind_var_clean(dataframe,merge_column,test_column,selection_column,output_name):
2      # remove all rows that aggregates data from SCHOOL_DSTRCT_CD
3      dataframe = dataframe.loc[dataframe[merge_column] != 'ALL']
4      dataframe['School District Code'] = dataframe[merge_column].astype(int)
5      dataframe = dataframe[test_column]
6      # remove all the rows with NaN
7      clean_data = dataframe.dropna()
8      # isolate the two columns that we want to test with group by
9      clean_data.groupby(['School District Code',selection_column]).mean()
10     clean_data = clean_data.reset_index(drop='True')
11     clean_data.to_csv(f'../../merge_ready_data/{output_name}.csv')
```

Analysis: Juptyer Notebooks:

```python
# merged_academic = merged DataFrame
# description = label for the independent variable
# selection_column = column where the labels of the independent variables are
# selection_column = values for the independent variable

def analysis(df, ind_label,ind_value):
    # import the dependencies
    import matplotlib.pyplot as plt
    from scipy import stats

    # build df based on school district code and independent variable label
    grouped = df.groupby(['School District Code', ind_label]).mean().reset_index()
    # create list of indepenent variables
    ind_var_list = grouped[ind_label].unique()
    # start for loop for each independent variable
    for ind_var in ind_var_list:
        ind_df = grouped.loc[grouped[ind_label] == ind_var]
        # check to see if there are actually data
        if ind_df[ind_value].mean()==0:
            continue
        ind_df = ind_df.assign(z_score = stats.zscore(ind_df[ind_value]))
        ind_df = ind_df.loc[(ind_df['z_score']<3) & (ind_df['z_score']>-3)]
        # build the regression function
        x_values = ind_df[ind_value]
        y_values = ind_df['Percent']
        (slope, intercept, rvalue, pvalue, stderr) = stats.linregress(x_values, y_values)
        regress_values = x_values * slope + intercept
        # print analysis
        print(f'regression function = f(x) = x*{slope.round(2)} + {intercept.round(2)}')
        print(f'p-value: {pvalue}')
        print(f'std error: {stderr}')
        ind_df.plot.scatter(x = ind_value, y = 'Percent', figsize = (8,8))
        plt.plot(x_values, regress_values, color = 'red')
        plt.title(f'{ind_var} vs college enrollment')
        plt.xlabel(ind_var)
        plt.ylabel('College Enrollment Percentage')
        ind_var = ind_var.replace("/", "&")
        plt.savefig(f'../images/{ind_var}.png')
        plt.show()

```

Analyze Graduation Analysis

```python
merged_df = pd.merge(dep_var_df, df, on="School District Code", how="inner")
merged_df = merged_df.dropna()
merged_df['PCT_analysis'] = merged_df['BEGIN_PCT'] + merged_df['BEGIN_PCT']
merged_df = merged_df.replace('Economics/Business/Free Enterprise','Economics & Business & Free Enterprise')
merged_df.columns
```

Press Ctrl+Alt+Enter to execute cell

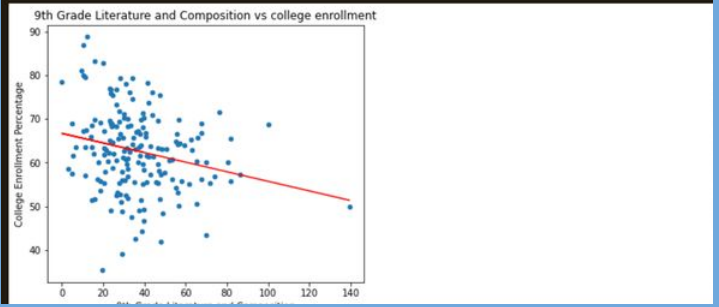
```
Index(['School Year', 'School District Code', 'School District Name',
       'School Code', 'School Name', 'Graduates', 'Postsecondary Institution',
       'Percent', 'Unnamed: 0', 'TEST_CMPNT_TYP_NM', 'NUM_TESTED_CNT',
       'BEGIN_CNT', 'DEVELOPING_CNT', 'PROFICIENT_CNT', 'DISTINGUISHED_CNT',
       'BEGIN_PCT', 'DEVELOPING_PCT', 'PROFICIENT_PCT', 'DISTINGUISHED_PCT',
       'PCT_analysis'],
      dtype='object')
```

```python
# merged_academic = merged DataFrame
# description = label for the independent variable
# selection_column = column where the labels of the independent variables are
# selection_column = values for the independent variable

merged_academic = merged_df
selection_column = 'TEST_CMPNT_TYP_NM'
ind_var_label = 'PCT_analysis'
```

```python
analysis(merged_academic,selection_column,ind_var_label)
```

```
regression function = f(x) = x*-0.11 + 66.72
p-value: 0.0017624595514645625
```



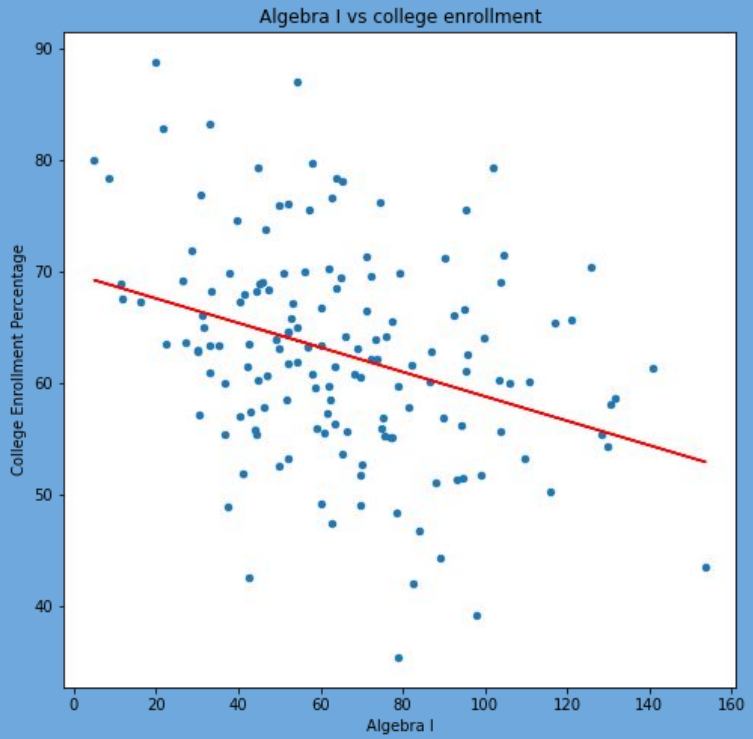
Analysis and Conclusion 1:
Allocate More Funding to 9th Grade Instructors

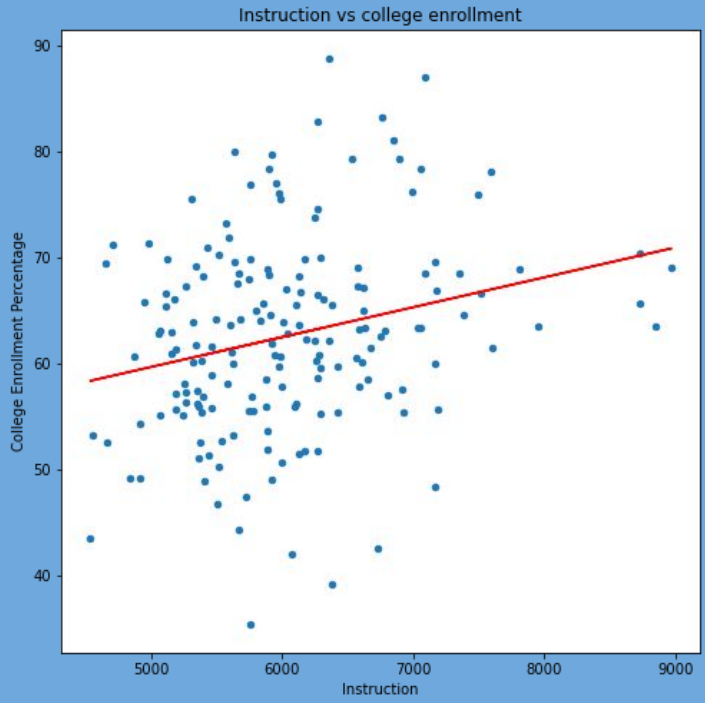
- Based on analysis on impact of college enrollment based on classes offered in high school, the most impactful, by p-value, classes were 9th grade Literature, Algebra I, Geometry. We realize those are the most common subjects for 9th grade students in high school.
- We evaluated revenue performance, specifically expenditure per student and found that increasing expenditure on teachers had a higher impact, by p-value, to college enrollment.

In conclusion, we believe spending more money on instructors teaching 9th grade classes will improve the chance of students enrollment into college after graduation most effectively. We think it's because 9th grade courses lay a foundation for expectation of students and their willingness to learn for the rest of the high school years.

* CSV's used: Revenue & Expenditures, Ga Milestones EOC Assessment (by grade)

Analysis and Conclusion 1 - Graphs:
Allocate More Funding to 9th Grade Instructors





Analysis and Conclusion 2:

Implement Specific Programs to Aid Homeless Students

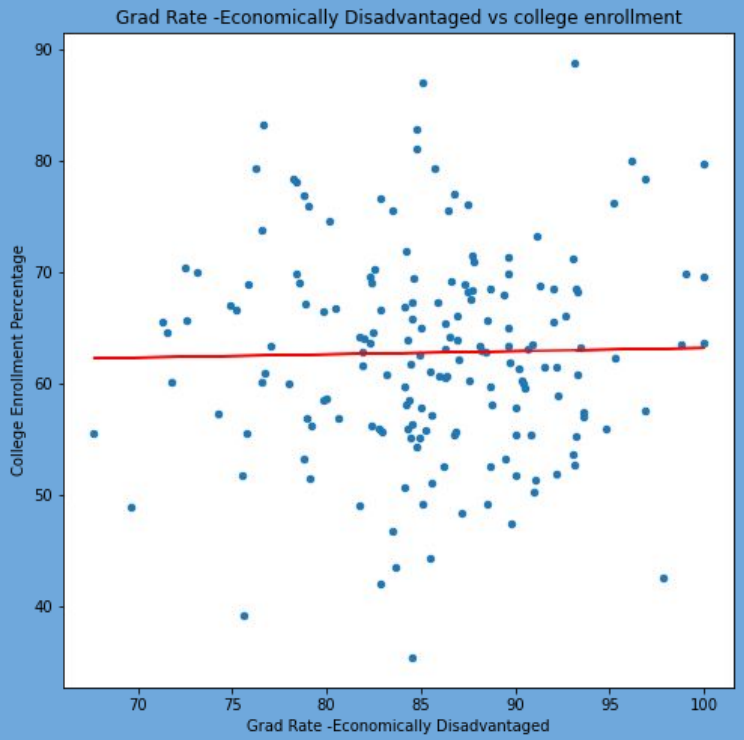
- School districts with higher percent of graduated homeless students had a lower college enrollment percentage.
- Analysis also showed economically disadvantaged students did not have a negative impact in college enrollment.

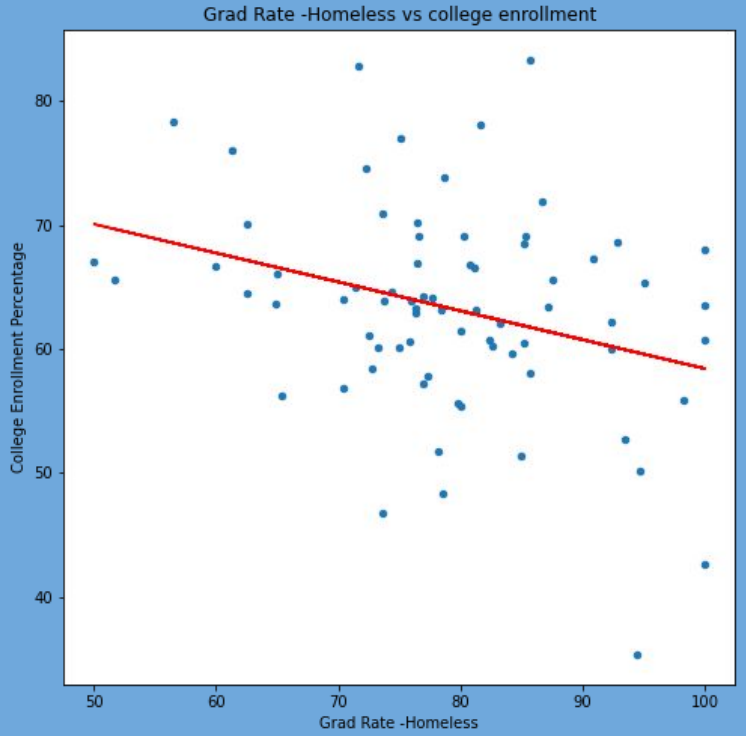
Considering the lack of significance of the economically disadvantaged students based on graduation rate, we believe the current programs are helping our low income students go to college. We should model our homeless programs to the economically disadvantaged students programs as well as implementing specific programs to aid homeless students.

*CSV's used: Graduation Rate (4-Year Cohort)

Analysis and Conclusion 2 - Graphs:

Implement Specific Programs to Aid Homeless Students





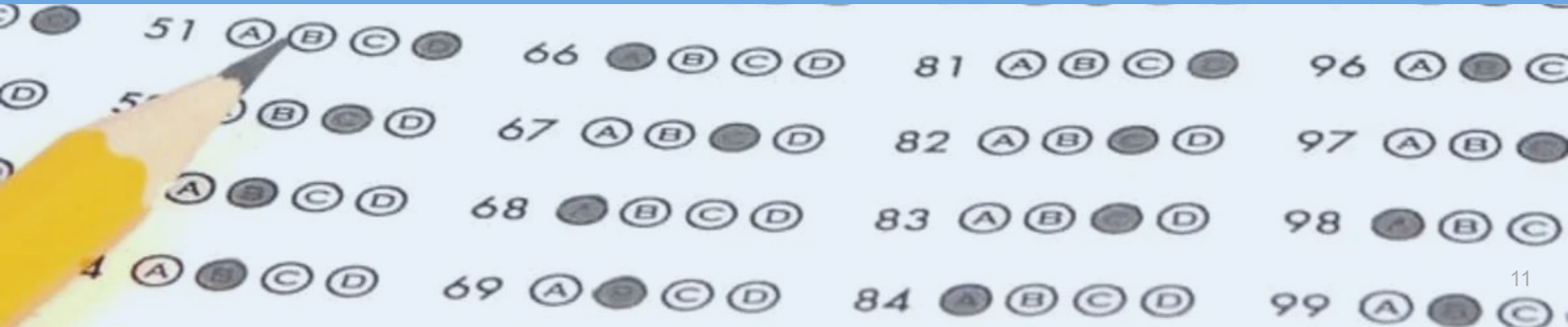
Analysis and Conclusion 3:

Schools Should Offer an SAT Prep Class/ Provide Support to Students

- As one would probably expect, a higher SAT/ACT score has a strong positive correlation with college enrollment
- However, we know it isn't possible to achieve a high score without test prep, which isn't always accessible to all students

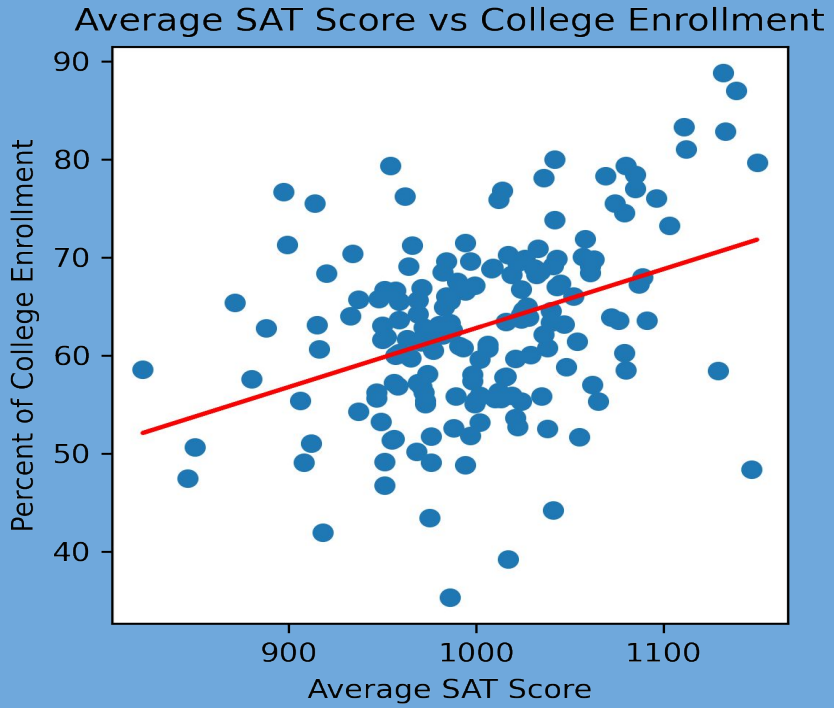
In conclusion, mandating an SAT/ACT prep course to all students regardless of income/demographic status should result in an increase in college enrollment

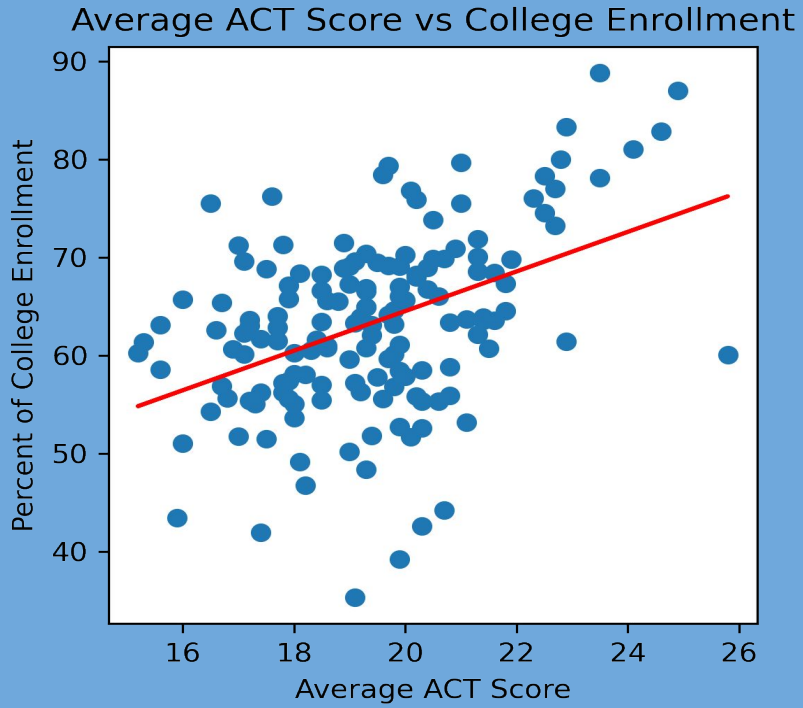
*CSV's used: SAT(Recent), ACT(Recent)



Analysis and Conclusion 3- Graphs:

Schools Should Offer an SAT Prep Class/ Provide Support to Students





Implications of Our Findings:

How Georgia School Districts can Improve College Enrollment Rates

- Spend more funds on hiring qualified 9th grade teachers
- Emphasize importance of success in 9th grade courses
- Implement specific programs to aid homeless students
- Offer SAT/ACT prep to all students in schools

Future Research:

- Build relationship between property value & college enrollment
- Do similar analysis for middle school success & impact on 9th grade success
- Noticed a strong relationship between English speaking proficiency and college enrollment rate


Thank You!
Any Questions?