

Texas Education

(Predicting the Percentage of Students Who Will Graduate College Within Four Years, Based on the Features of the School District They Attended High School in)

Capstone 1 Project

Data Collection:

Texas' Major Regions -The Focus of This Study

In this study, 2011 - 2017 data was collected from the following two resources:

-Texas Education Agency ("TEA"):

<https://tea.texas.gov/>

-Texas Education Reports ("TER"):

<https://www.texaseducationinfo.org/>

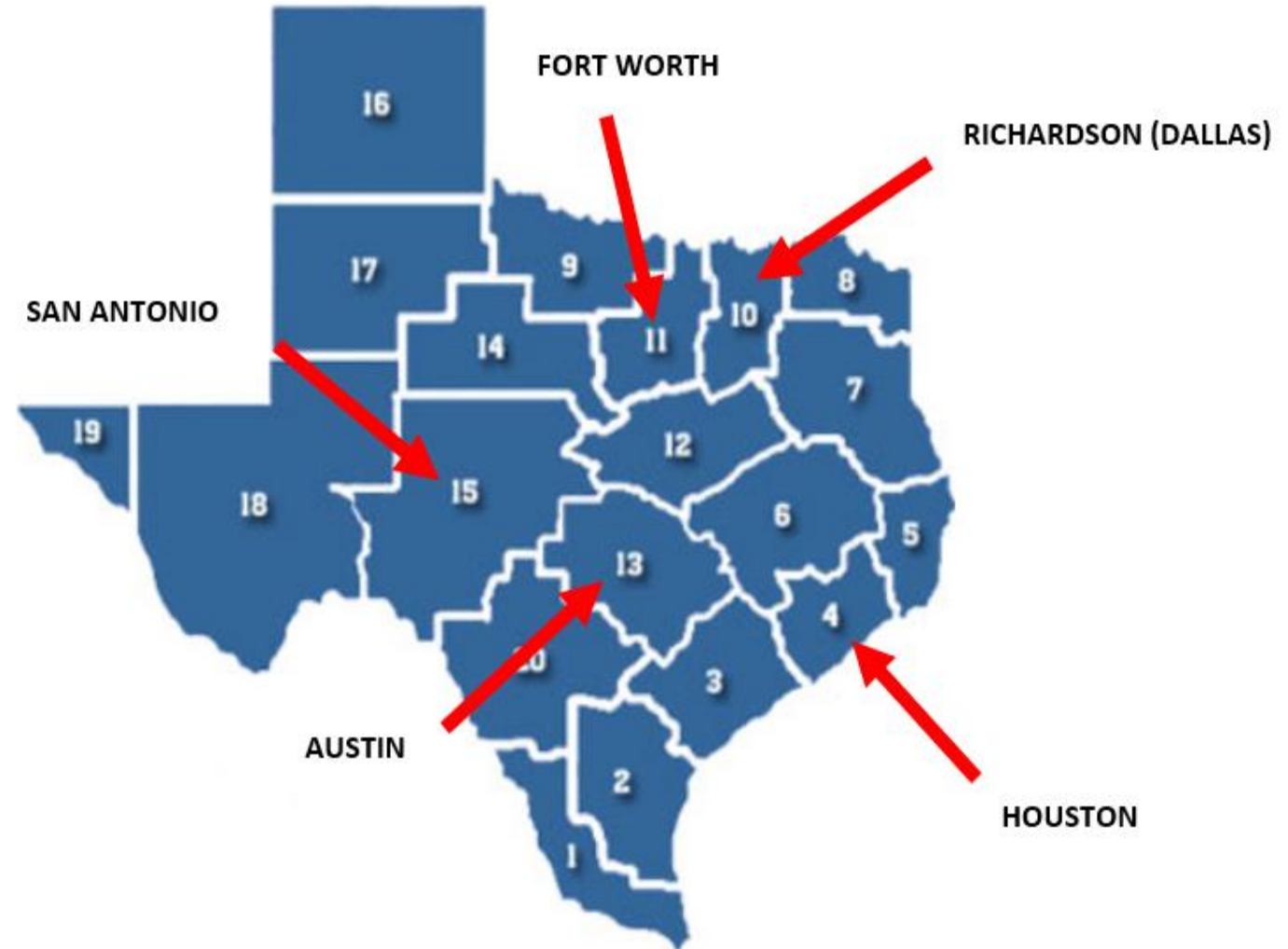
The data was then filtered to focus public school districts within the major regions of Texas.

The college enrollment and graduation data only considers Texas colleges.

<u>Major Region</u>	<u>Population (Millions)</u>
Houston	2.3
San Antonio	1.5
Richardson (Dallas)	1.3
Austin	0.95
Fort Worth	0.87

Texas' Major Regions

- Here's how the Texas Education Agency splits up its educational regions for reporting data
- The major regions represent areas with the most economic opportunity, making them prime locations for families and talented educators.



“Which School Districts are Proven to Contain a Higher Percentage of its Graduates Earn a College Degree Within Four Years?”

- The above question was presented by clients (parents) considering a move to or within the major regions of Texas. The clients would like their child to one day attend a college in Texas.
- They are looking to avoid a situation in which their child takes more time to earn their degree (more tuition money spent) or even fails out (worst case scenario with no return on investment).
- In strictly focusing on Texas colleges, the clients are also looking to avoid expensive out-of-state tuition.



Importance Behind the Question & Providing a Solution:

The clients have expressed that they simply want their child to live a more comfortable life after earning a college degree and receiving a quality education.

- I approached the problem presented to me by collecting historical data on the percentage of students who earned a college degree within four years' time after graduating high school from a particular school district. For each school district, their historical features were also collected to measure their influence on the resulting college graduation percentage.
- Utilizing this historical data, I aimed to build a predictive model to estimate the percentage of students going to college (from a specific school district) that will graduate within four years.

College Admissions Tests

SAT

- Highest Possible Score is a 1600

$$\text{SAT(Total)} = \text{Math (800)} + \text{Reading/Writing (800)}$$

ACT

- Highest Possible Score is a 36

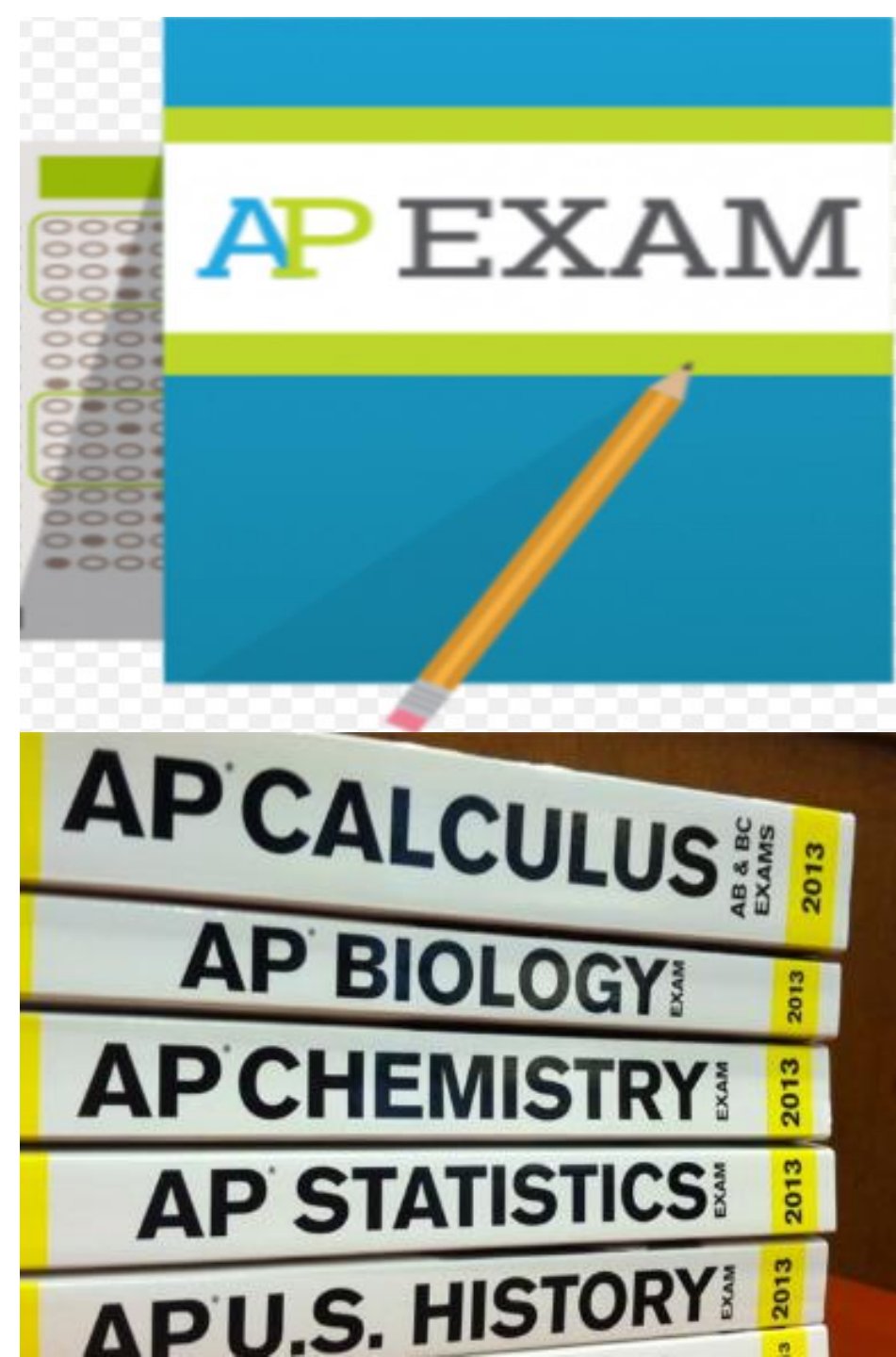
$$\begin{aligned} \text{ACT (Composite)} = \\ (\text{Math (36)} + \text{Reading (36)} + \text{English (36)} + \text{Science (36)}) \\ / 4 \end{aligned}$$



AP Exams

At the end of the school year, students enrolled in AP classes have the chance to earn college credit through AP Exams!

- Most colleges nationwide offer college credit for qualifying AP exam scores. College courses can cost thousands of dollars, but if you take and pass an AP test, you're only spending roughly \$93.
- AP exams are scored from one to five, with five being the highest score. Colleges will accept a minimum exam score for it to transfer to college credit.
- For the purpose of this data analysis, we will use the benchmark score of a 3 (most colleges accept this score) as an indicator of how well students in a district test.



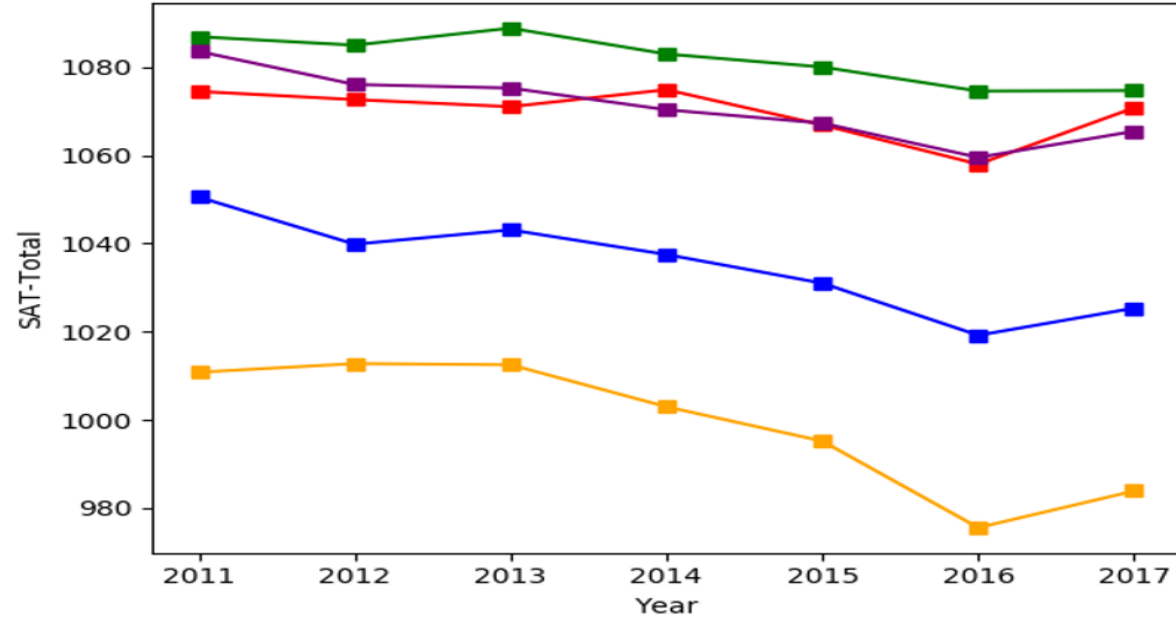
Wealth Per Average Daily Attendance ("Wealth/ADA")

- Wealth/ADA is simply the property value of each school district divided by its average daily attendance.
- The property value comes from the Texas state comptroller and is the basis for each school district's local property tax collections.

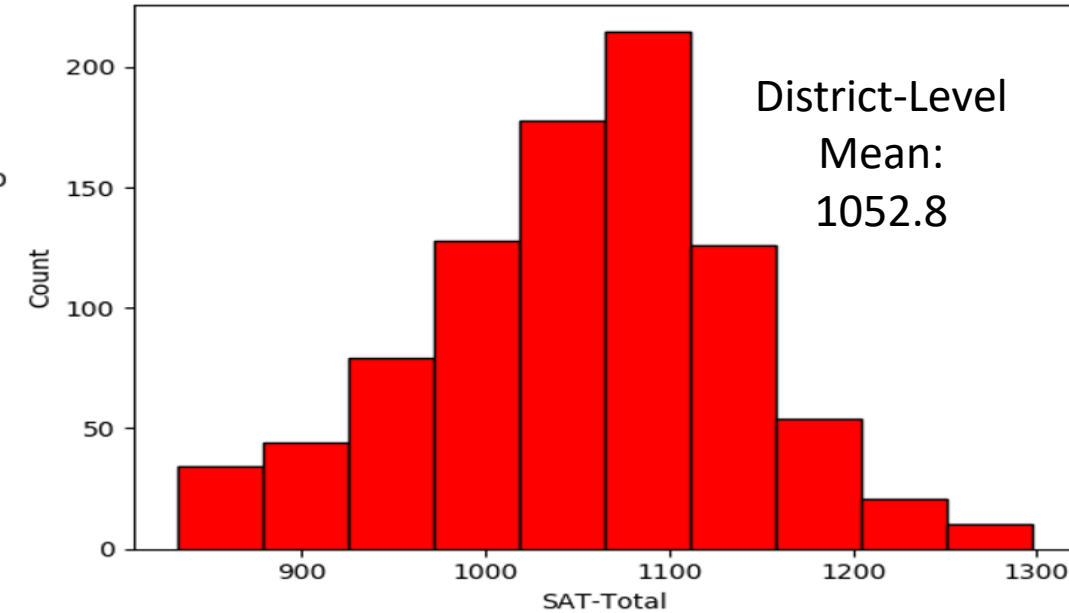


Data Analysis: (SAT & ACT)

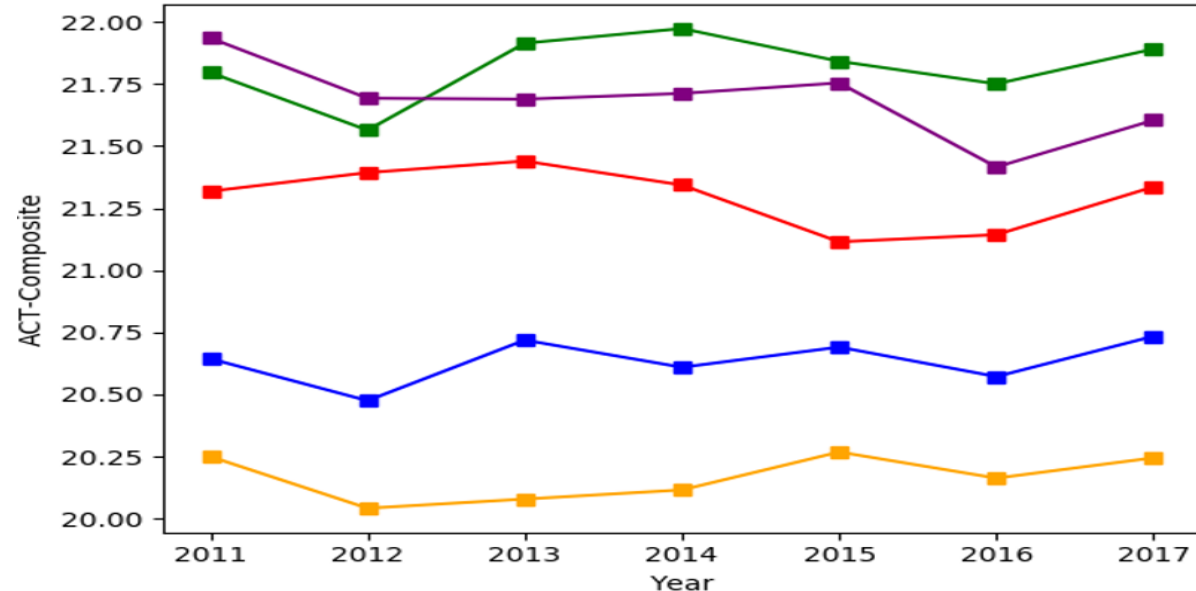
Regional Average SAT-Total (2011 - 2017)



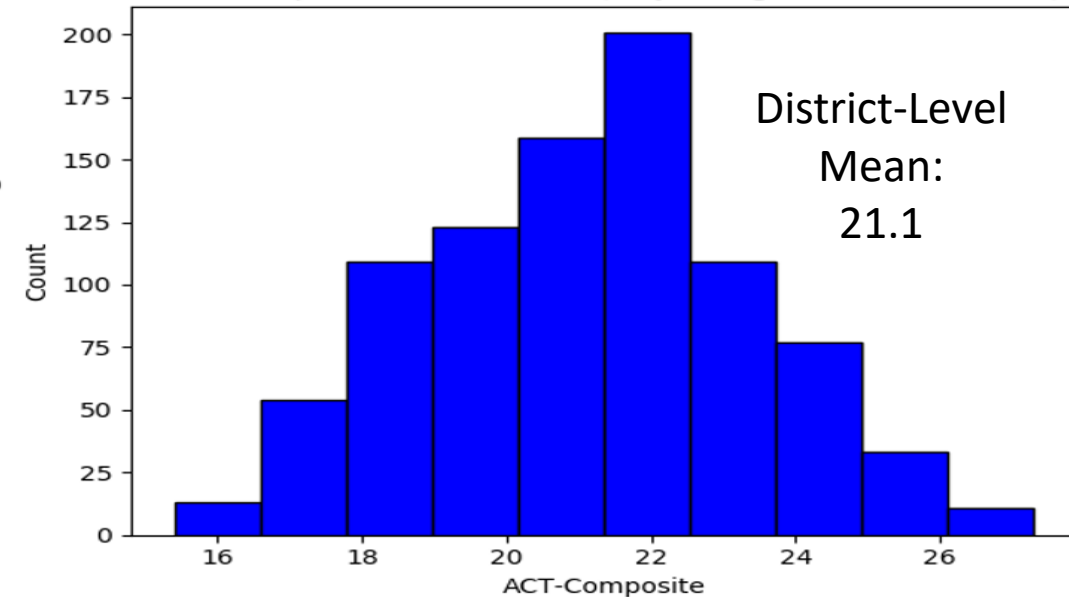
SAT-Total Distribution (Major Regions: 2011 - 2017)



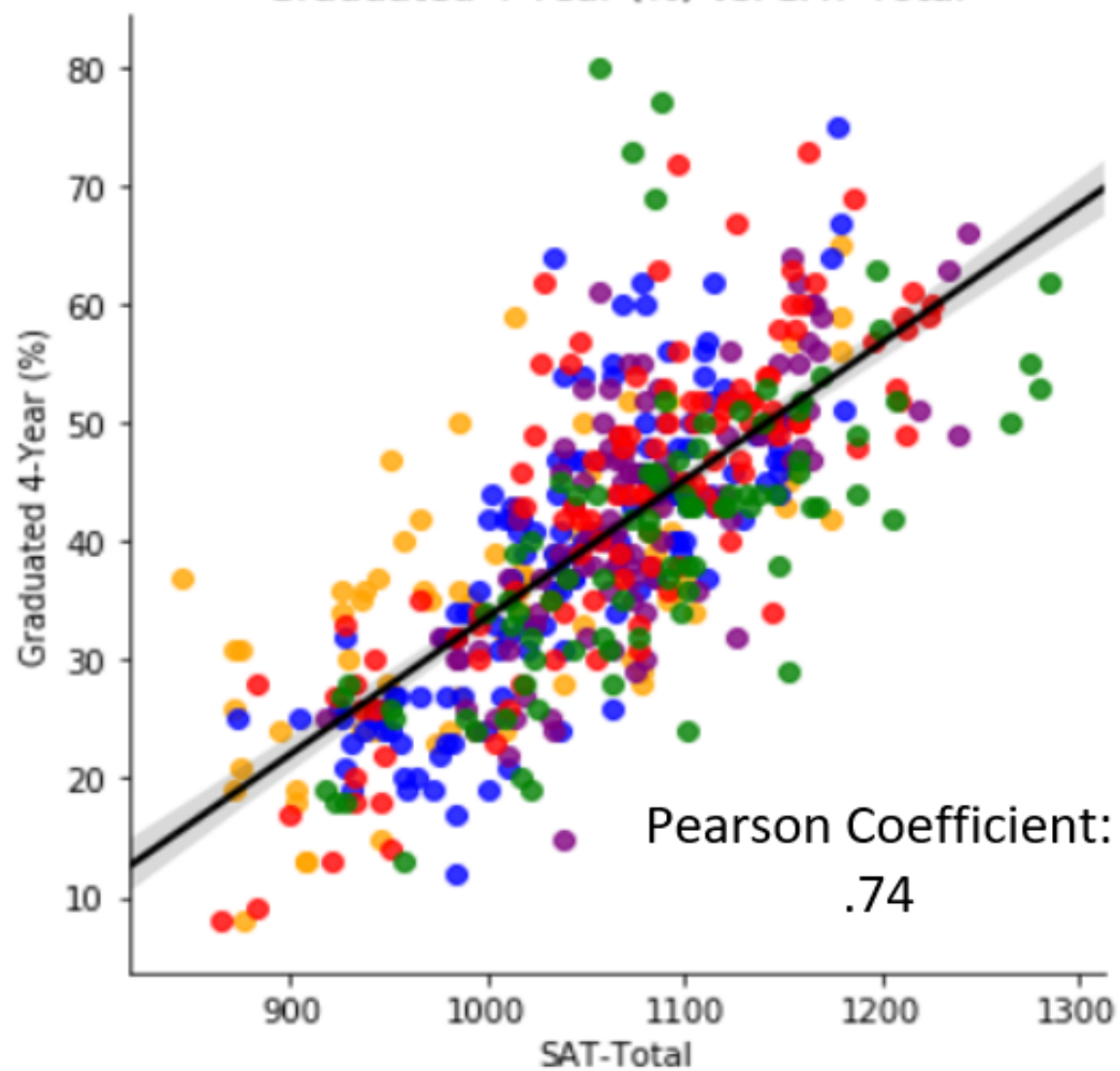
Regional Average ACT-Composite (2011 - 2017)



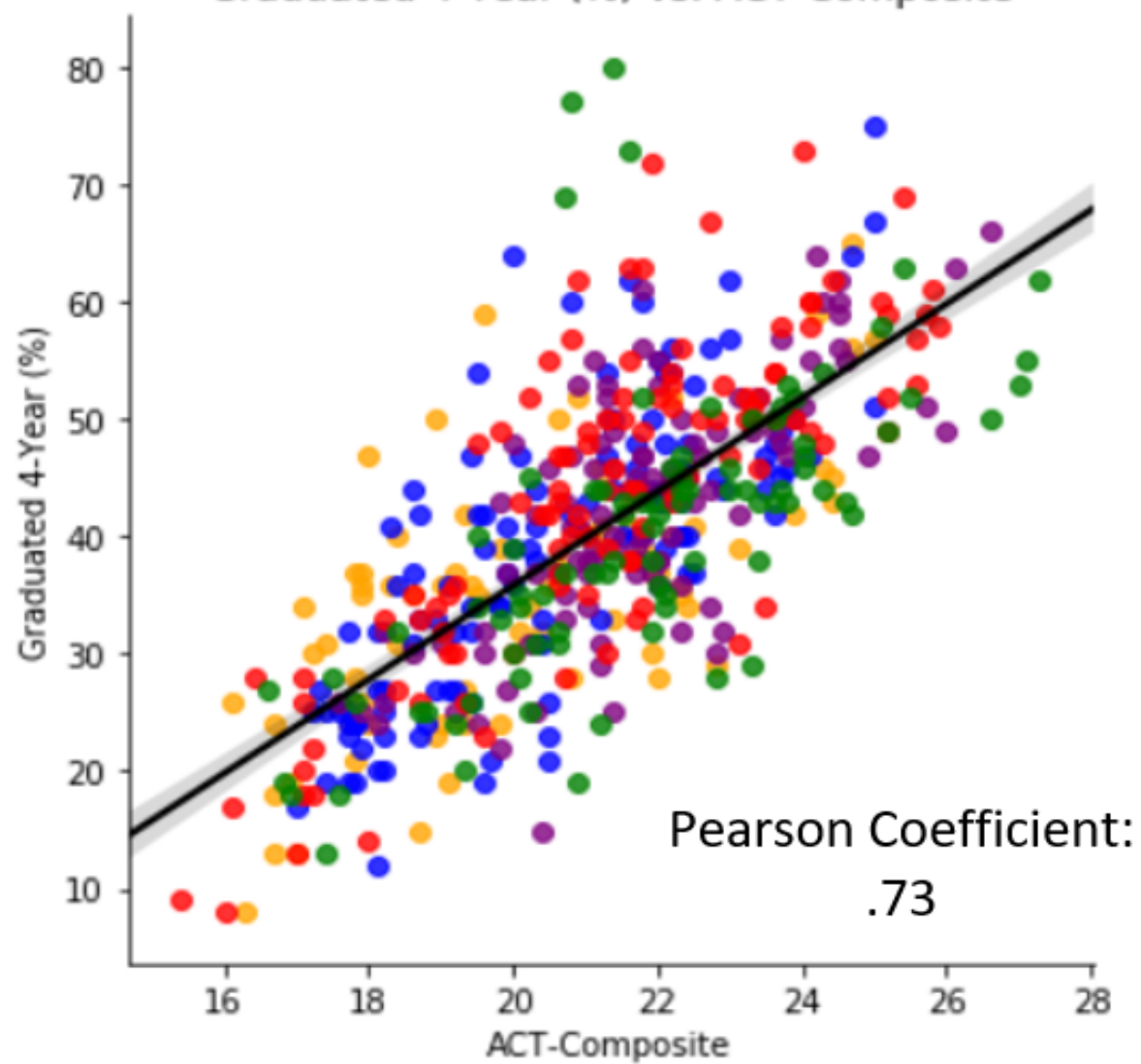
ACT-Composite Distribution (Major Regions: 2011 - 2017)



Graduated 4-Year (%) vs. SAT-Total



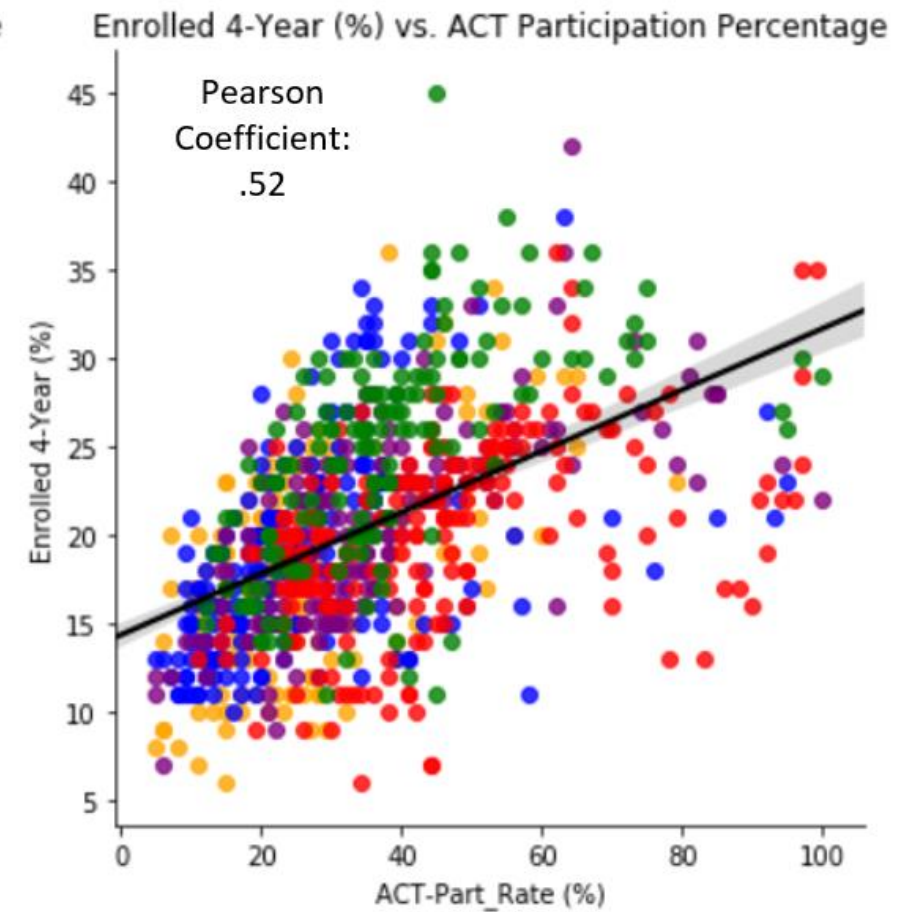
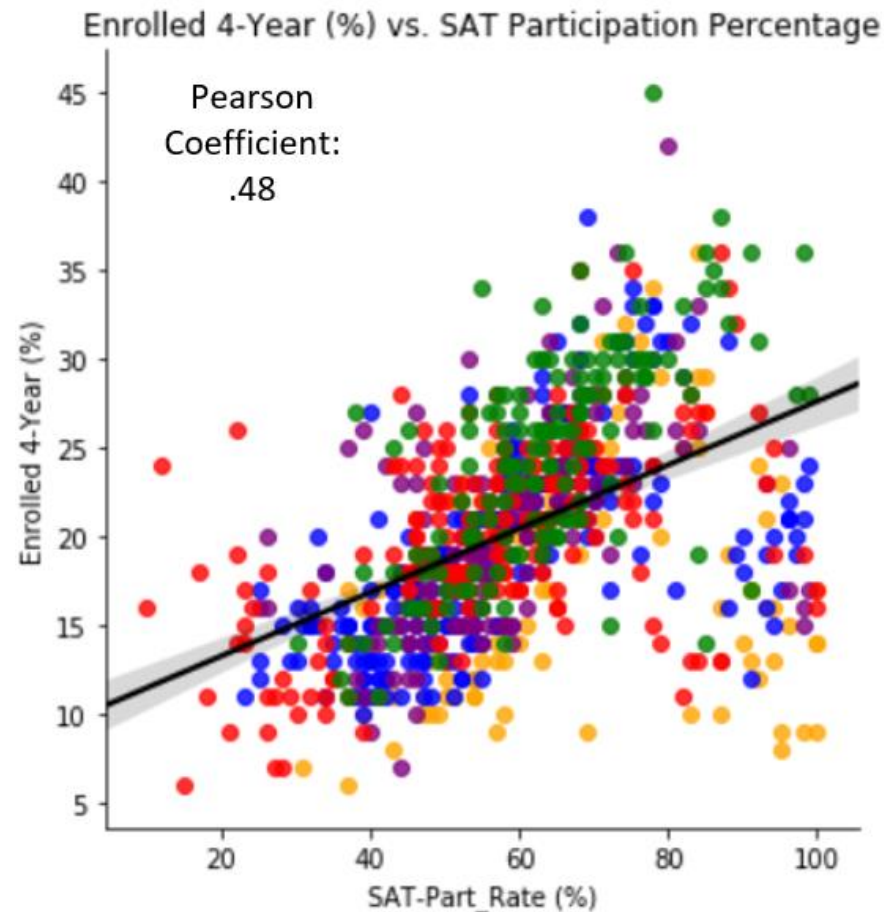
Graduated 4-Year (%) vs. ACT-Composite



Data Analysis: (SAT/ACT Participation %)

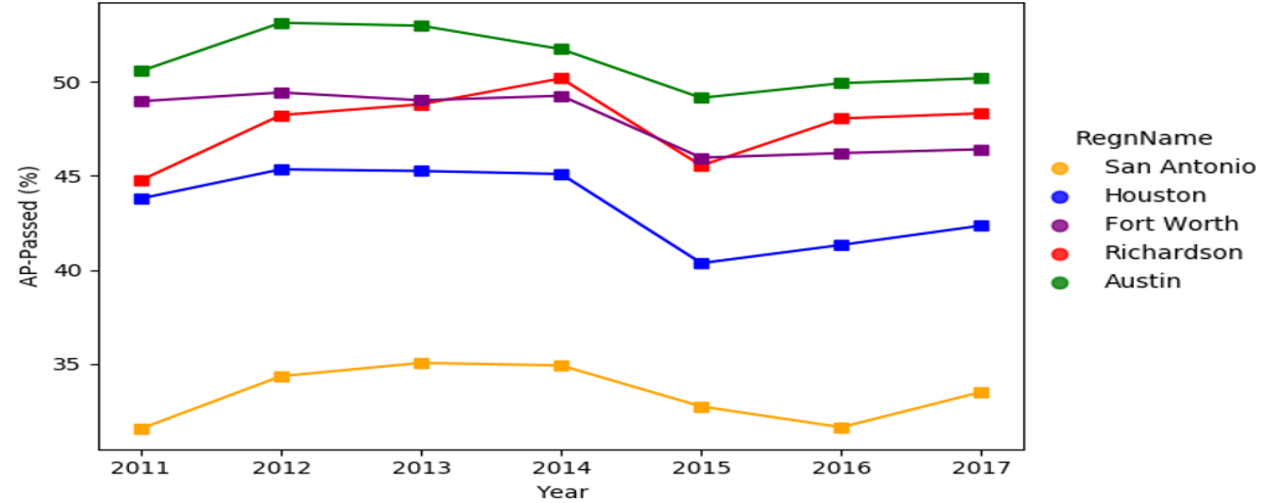
- It appears that the classes of 2011 – 2017 all appeared to favor taking the SAT. Why is this?
- Participation % for both tests contained similar positive correlations with college enrollment %

Year	2011	2012	2013	2014	2015	2016	2017
ACT-Part_Rate	32.007874	32.937008	32.346457	33.771654	35.204724	37.826772	37.409449
SAT-Part_Rate	61.866142	57.771654	56.511811	58.559055	59.039370	59.755906	62.267717

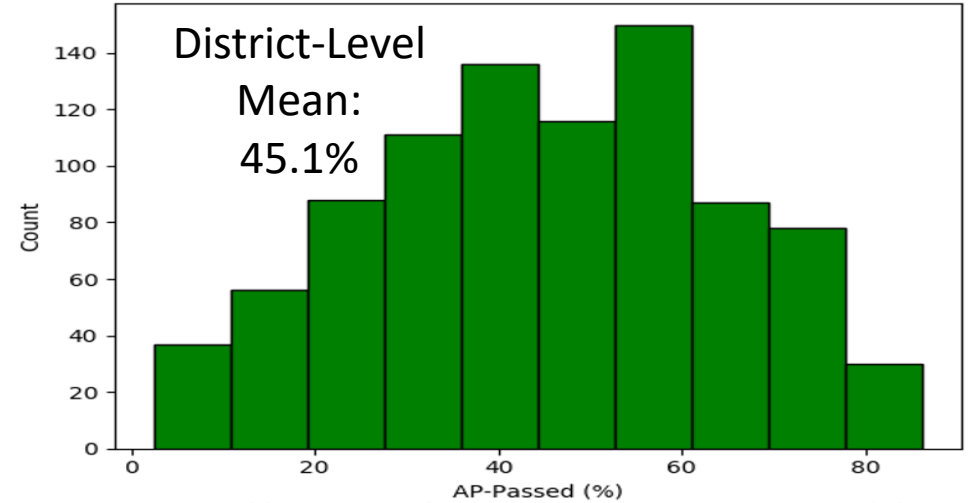


Data Analysis: (AP- Passing % and AP-Exams/Student)

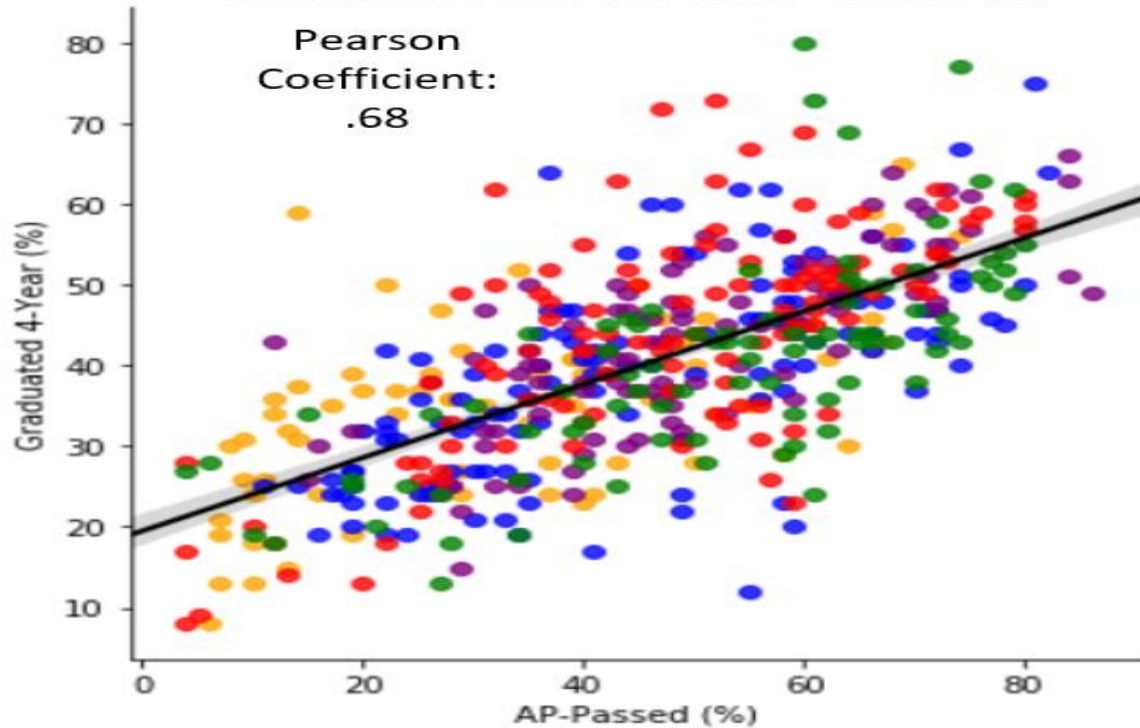
Regional Average Percentage of Passed AP Exams (2011 - 2017)



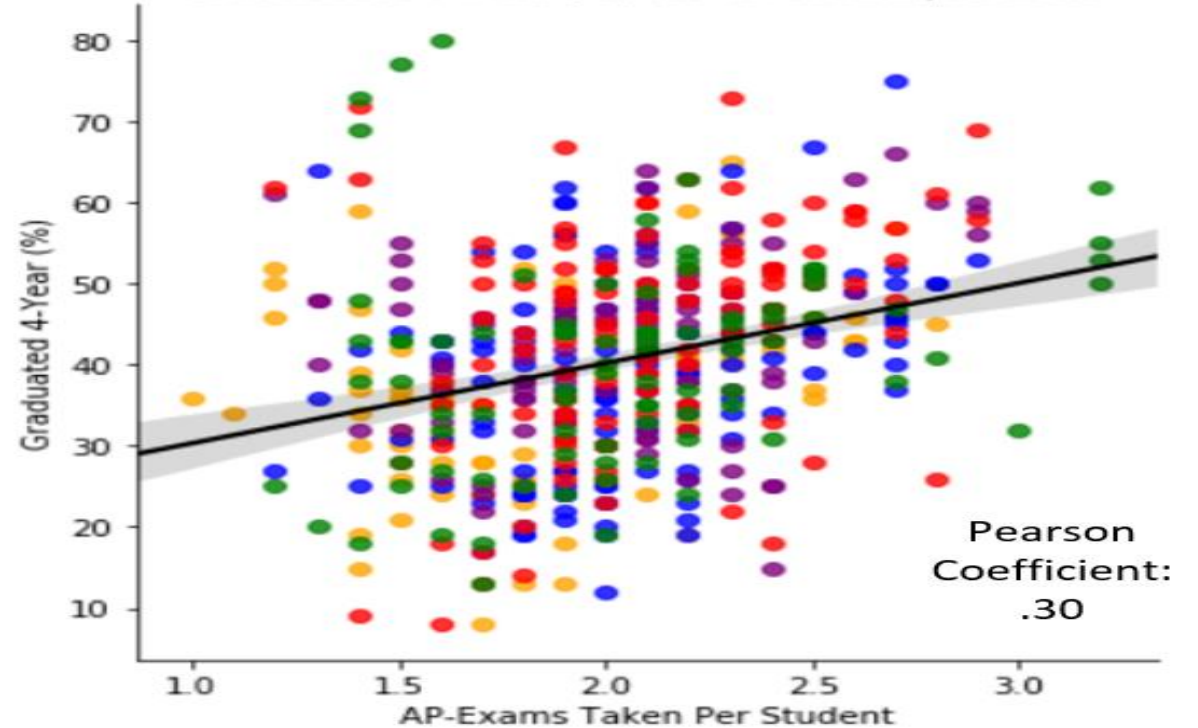
AP-Passed (%) Distribution (Major Regions: 2011 - 2017)



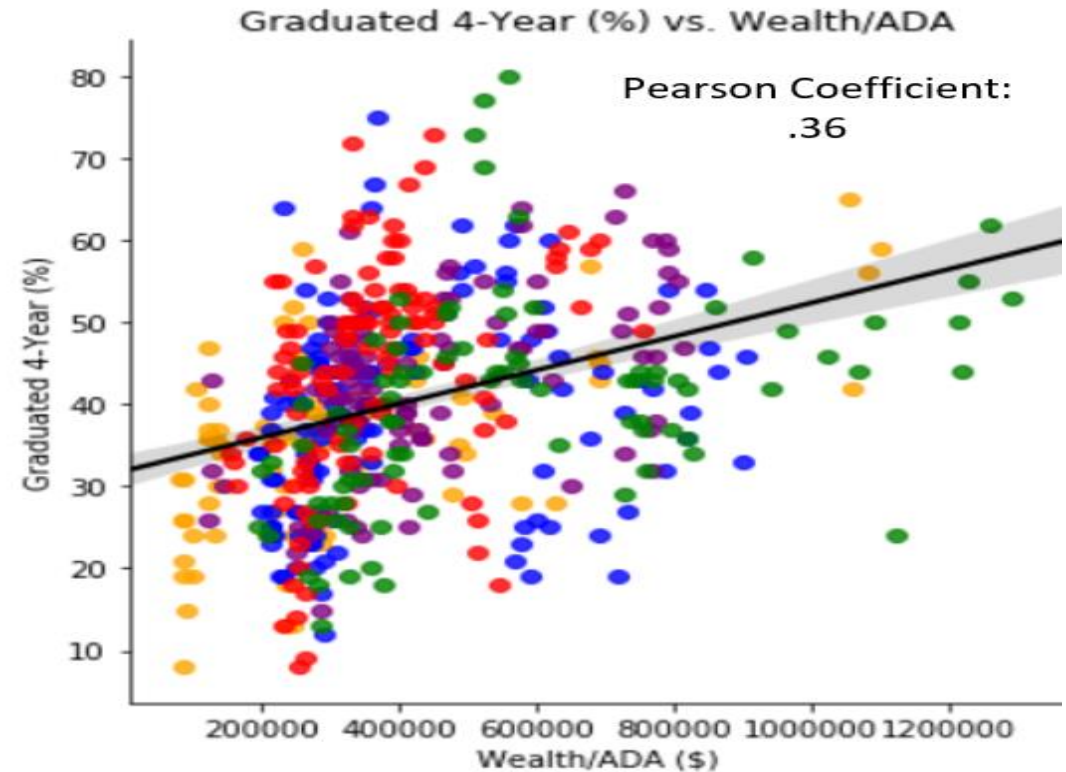
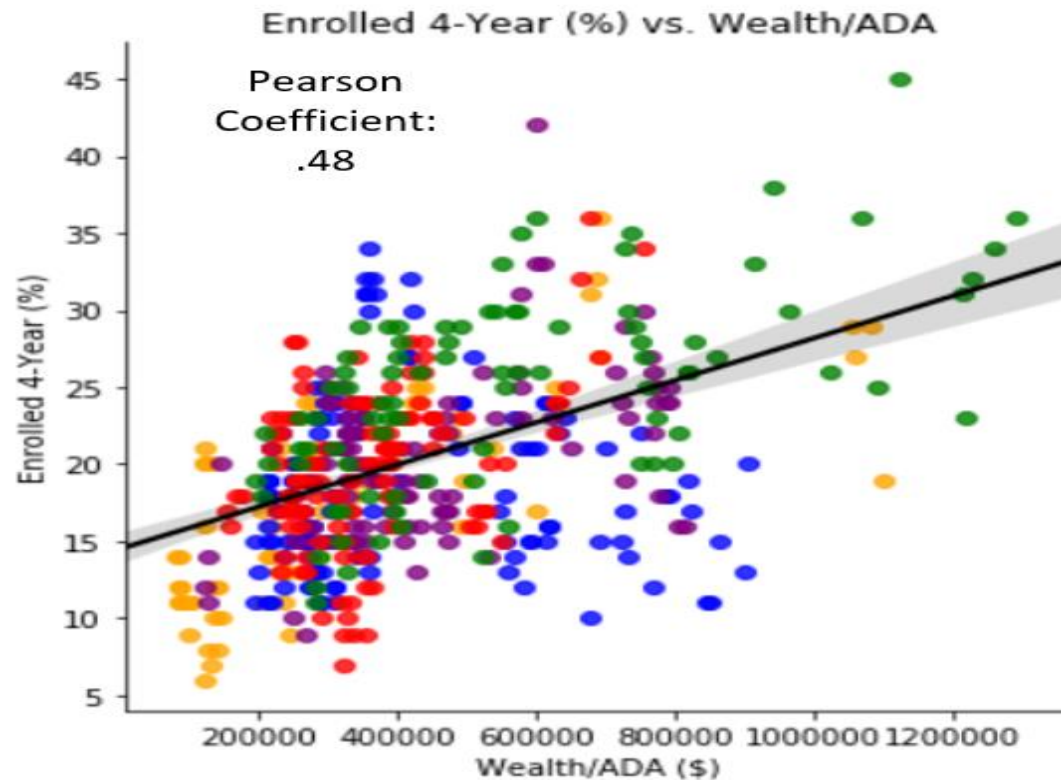
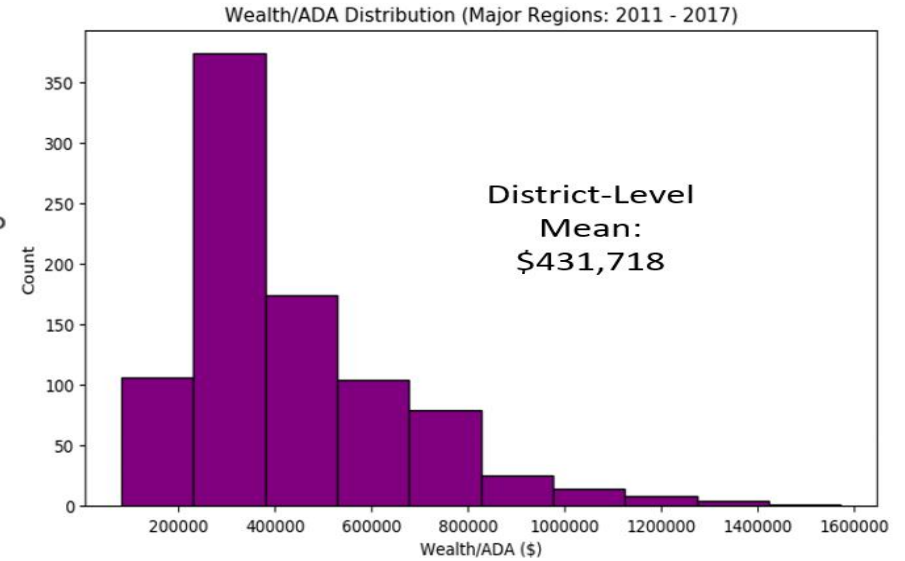
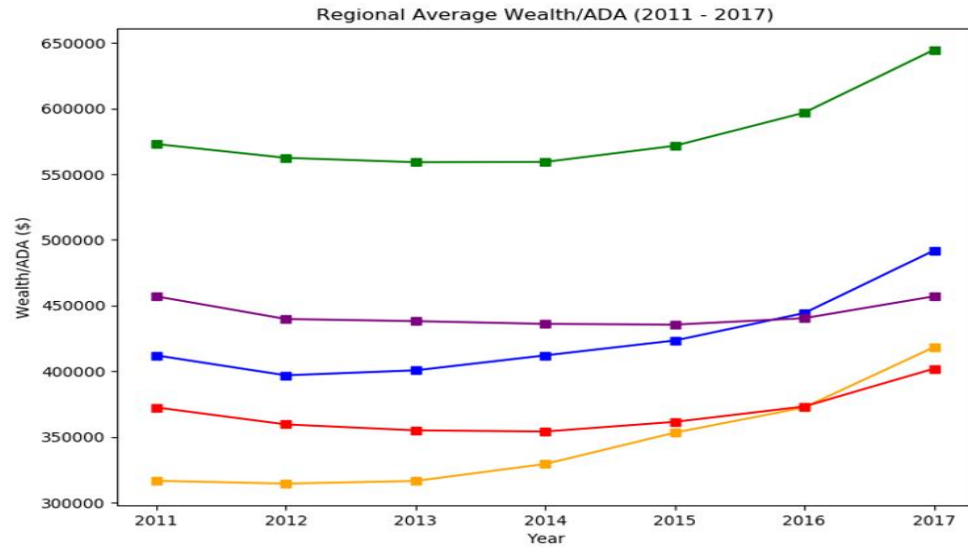
Graduated 4-Year (%) vs. AP-Passed (%)



Graduated 4-Year (%) vs. AP-Exams/Student

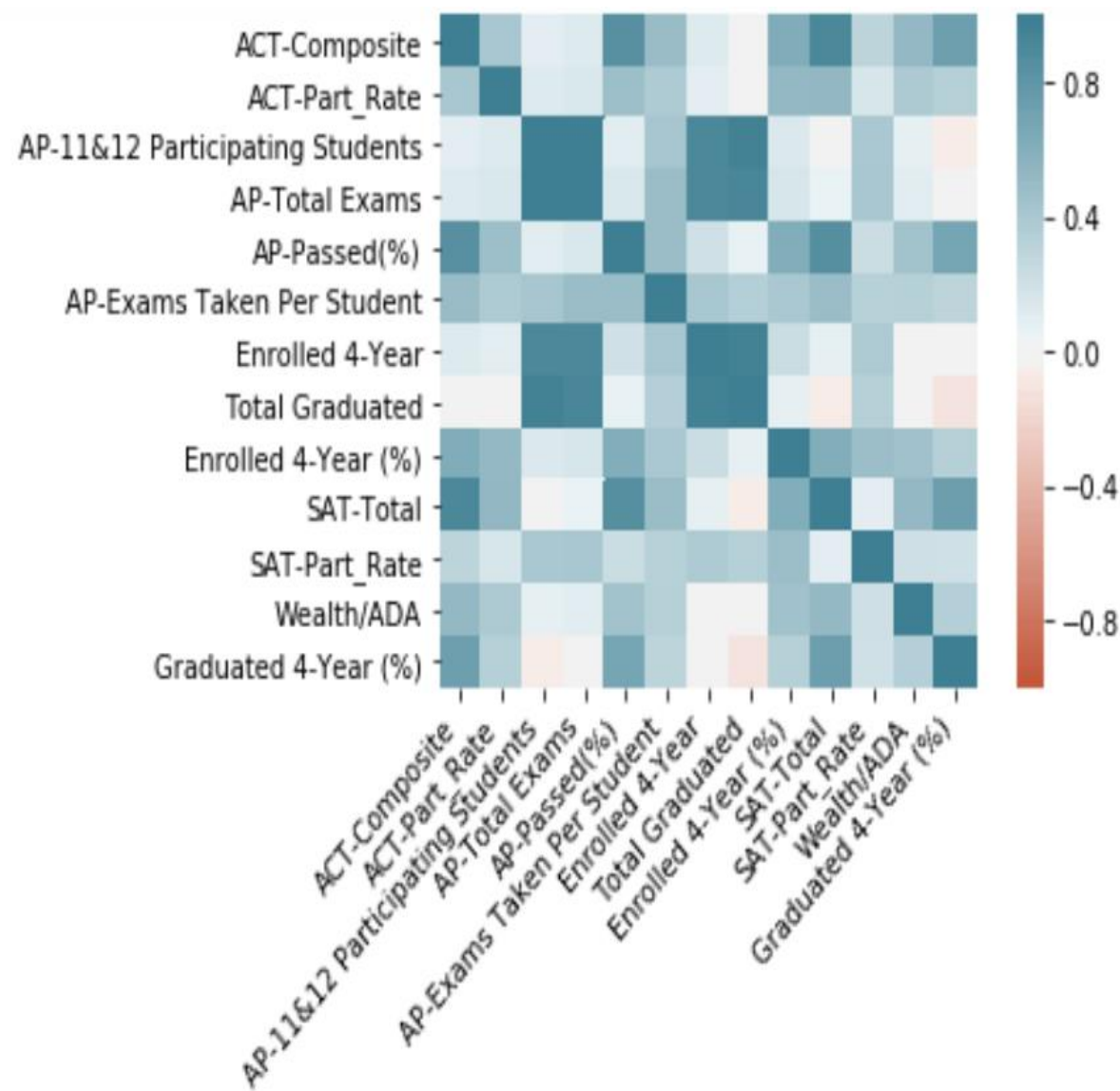


Data Analysis: (Wealth/ADA's Influence on College Enrollment (%) Versus College Graduation (%))

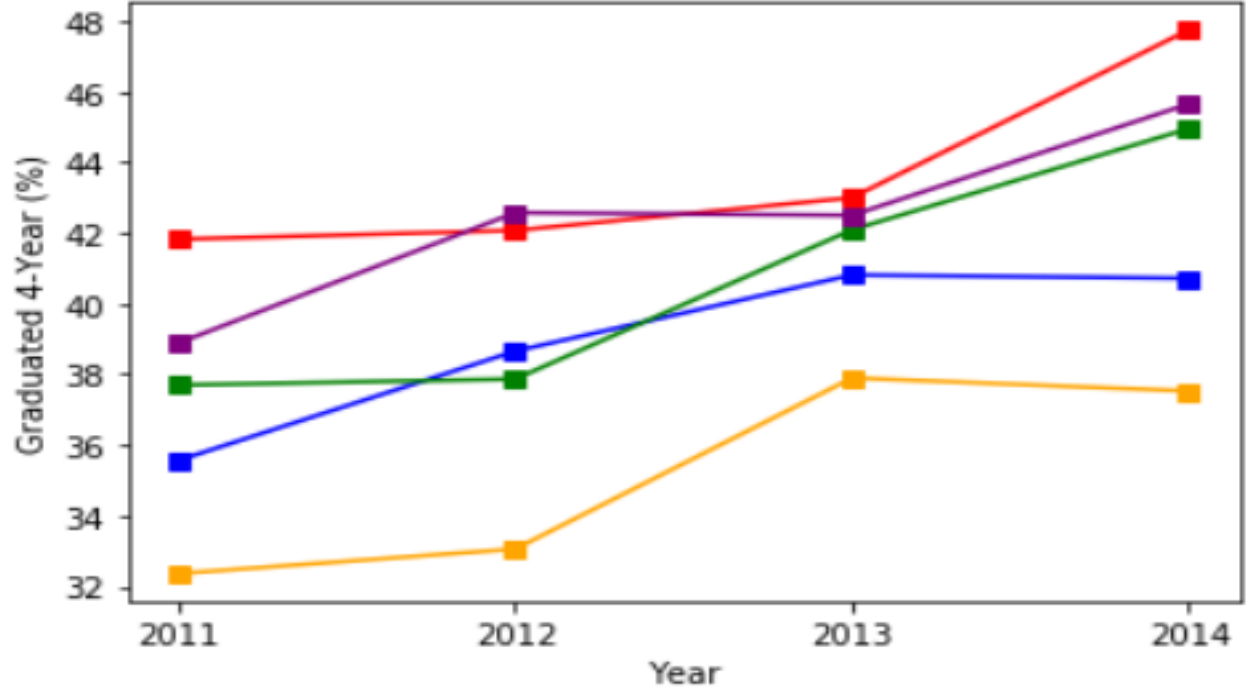


Data Analysis: *Historical Percentage of Students Earning a College Degree Within Four Years*

Correlation Heatmap:

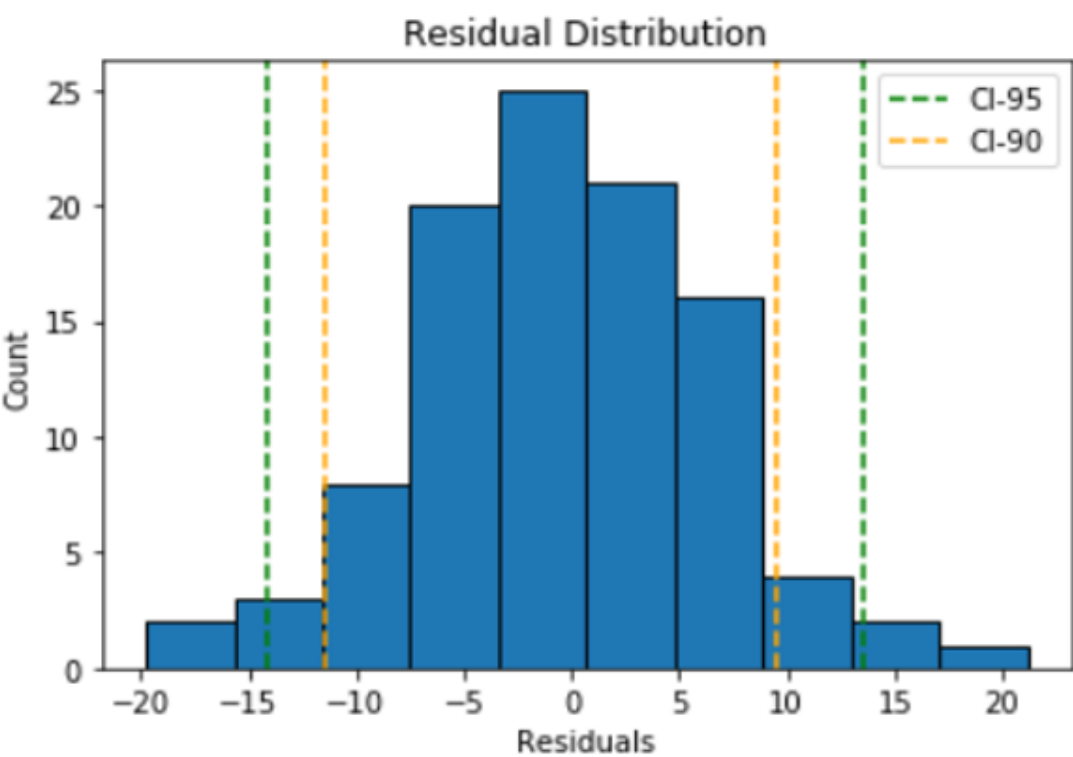
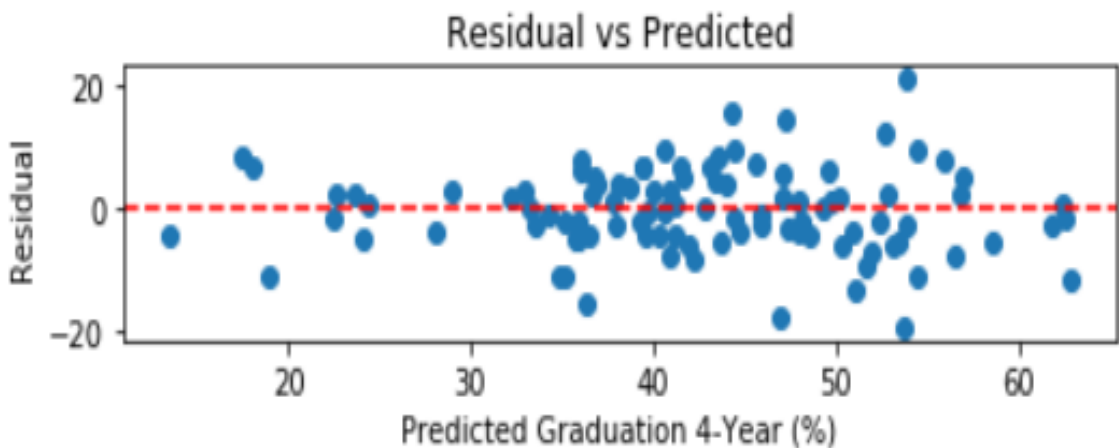


Regional Average College Graduation Percentage (2011 - 2014)



Graduated 4-Year (%)	
RegnName	
Richardson	43.660714
Fort Worth	42.404303
Austin	40.617418
Houston	38.937500
San Antonio	35.245899

Model 1: Linear Regression

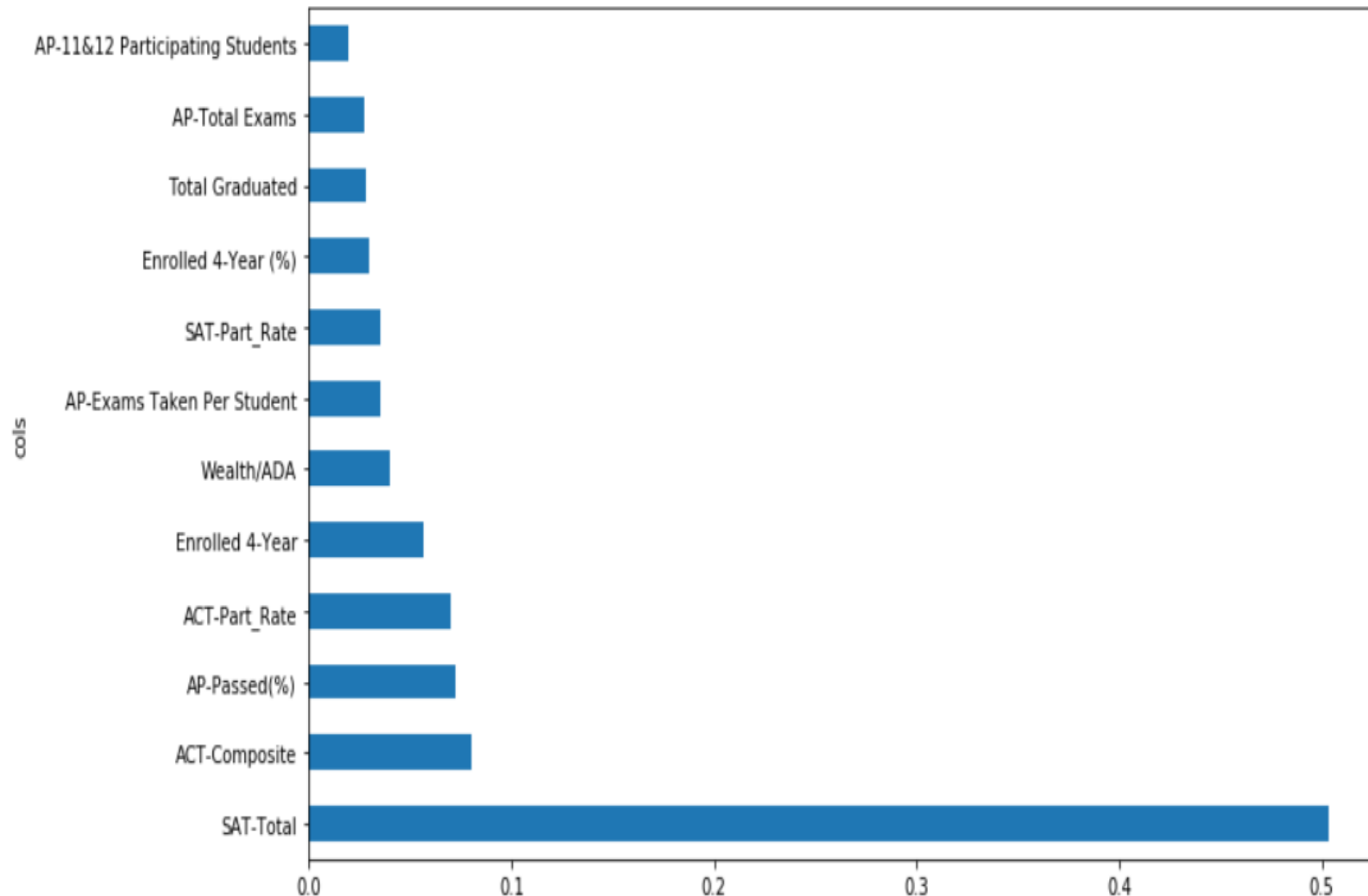


	RMSE	R2	MAPE (%)
Training	8.06	0.59	17.96
Testing	6.92	0.67	15.07

CI	lower bound	upper bound	CI Range
90	-11.387778	9.553608	20.941385
95	-14.172749	13.498170	27.670919

Model 2: Random Forest Regressor (No HyperParameter Tuning)

Feature Importance Plot:



OOB Score: 0.620

Testing RMSE: 6.697

Testing R2: 0.690

Testing MAPE: 15.537

- Decided to drop bottom four features for Model 3. Features were redundant and/or used to engineer other features

Model 3: Random Forest Regressor (HyperParameter Tuning & Feature Selection)

'n_estimators': [100, 200, 600, 1000]
'max_features': ['sqrt', 'log2', 0.5, None]
'min_samples_leaf': [5, 3, 2, 1]



Best Parameters:
'n_estimators': 1000
'max_features': 'log2'
'min_samples_leaf': 1

OOB Score:	0.629
Testing RMSE:	6.638
Testing R2:	0.696
Testing MAPE:	15.386

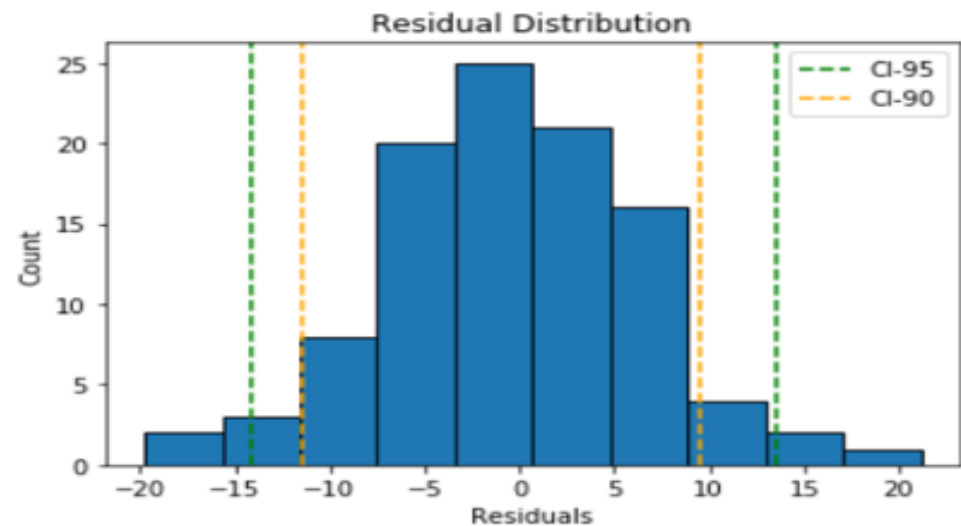
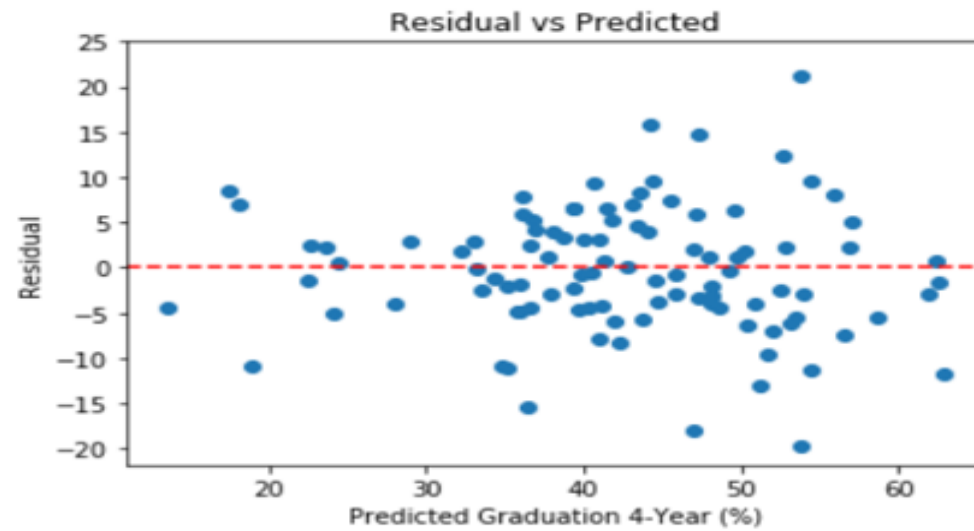
Model Selection – Performance Summary

Model	RMSE	R2	MAPE (%)	OOB
1	6.92	0.67	15.07	N/A
2	6.70	0.69	15.54	0.62
★ 3	6.64	0.70	15.39	0.63

Best Model: Model 3

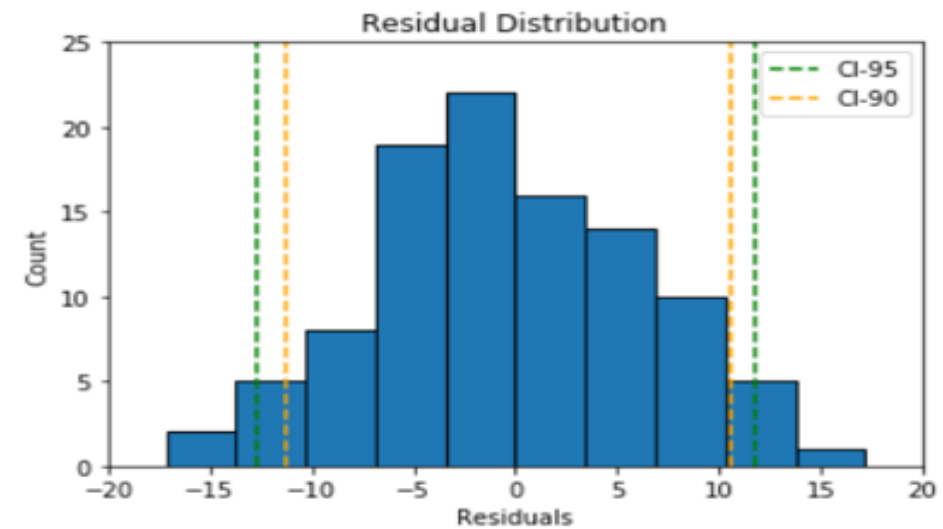
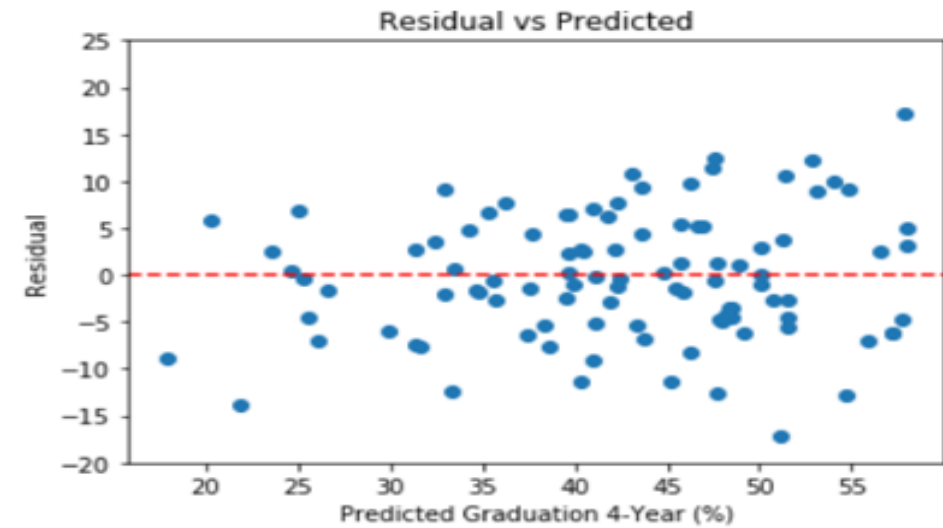
- While these performance results led me to favor Model 3, we'll also compare its “Residual vs Predicted” and residual distribution with the base model (Model 1) in the next slide.

Model 1:



CI	lower bound	upper bound	CI Range
90	-11.387778	9.553608	20.941385
95	-14.172749	13.498170	27.670919

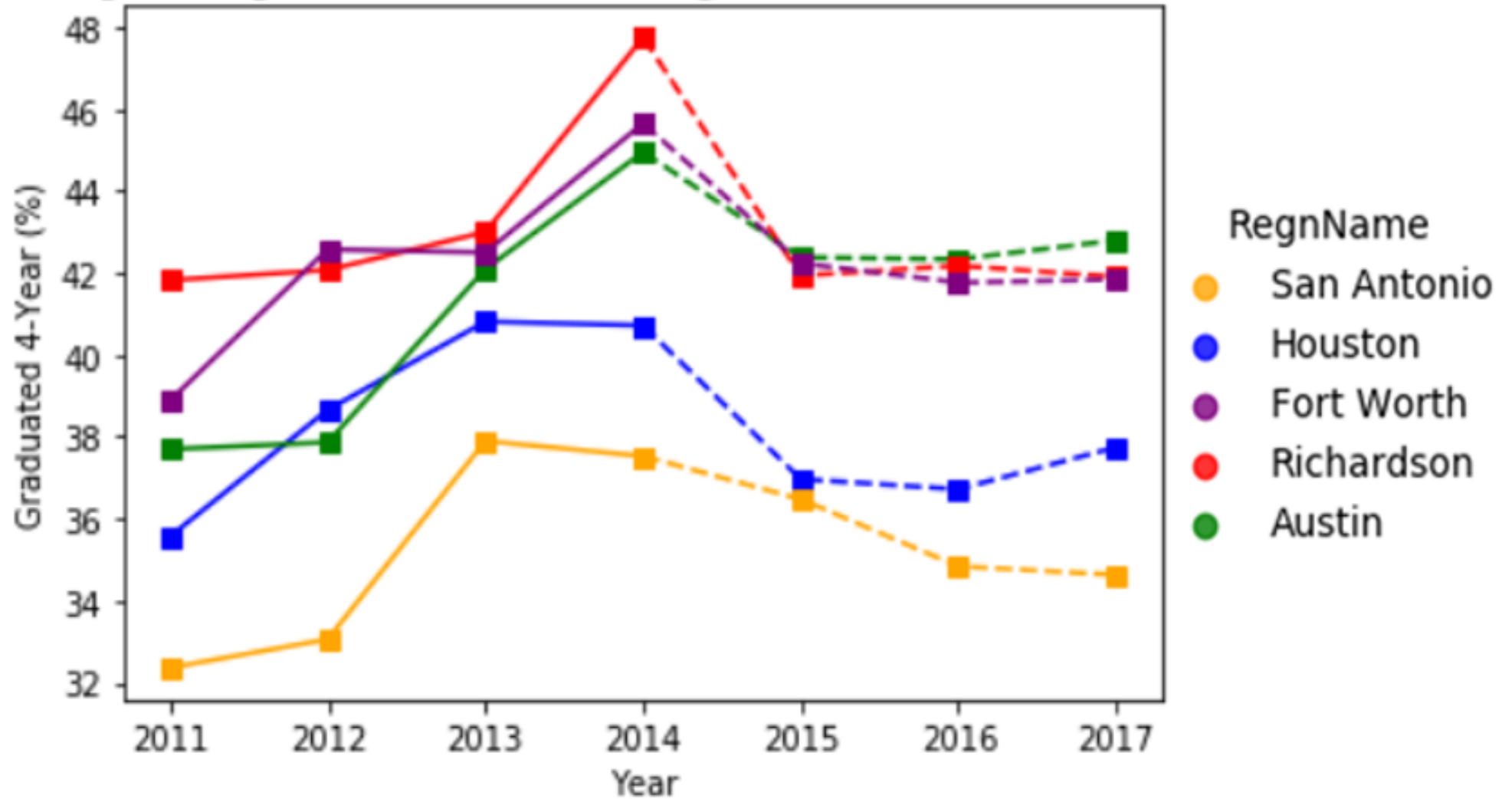
Model 3:



CI	lower bound	upper bound	CI Range
90	-11.254988	10.587250	21.842238
95	-12.706625	11.806625	24.513250

Predicting College Grad. % for High School Classes of 2015 - 2017

Regional Average College Graduation Percentage (Hist:2011 - 2014 + Pred:2015 - 2017)



Future Work

- When high school class of 2015 graduation percentage data come out, this will give me more historical data to train my model on, leading to better predicting power on unseen data.
- Neural Network
- Replicate project for other states (could lead to some interesting comparisons)
- Should data come out for students who were able to graduate within four years from a specific Texas college (Ex: “University of Texas” / “Texas A&M”) after graduating high school from a particular Texas school district, I could build a model for clients that already have that certain Texas college in mind for their child.

Recommendations to the Clients

Without any particular major region in mind, the client can simply take the approach of seeing who is predicted to average the best college grad. % for classes of 2011 – 2017.



DistName	RegnName	Graduated 4-Year (%)
LA GRANGE ISD	Austin	62.973895
FRIENDSWOOD ISD	Houston	61.648571
CARROLL ISD	Fort Worth	58.885714
LOVEJOY ISD	Richardson	58.777857
GRAPEVINE-COLLEYVILLE ISD	Fort Worth	58.410714

One could also narrow the search to district counts by region for those predicted to average college grad. % above 50%. From this standpoint, client would be advised to consider Richardson (Dallas) due to the number of options



RegnName	
Austin	4
Fort Worth	5
Houston	2
Richardson	10
San Antonio	1

Recommendations to the Clients (continued)

Some clients may prefer taking a look at the class of 2017 to view the predicted college grad. % for 2021. Here clients are taking more of a “what have you done for me lately” approach.

What region are you interested in? (Houston/San Antonio/Fort Worth/Richardson/Austin): Richardson
What high school class year of students are you interested in?: 2017
How many districts in your chosen region would you like to view?: 4

	DistName	RegnName	Class Year	Pred. 2021 College Graduation (%)
0	COPPELL ISD	Richardson	2017	60.280
1	PLANO ISD	Richardson	2017	57.730
2	CELINA ISD	Richardson	2017	57.710
3	FRISCO ISD	Richardson	2017	57.385

Recommendations to the Clients (continued)

- Many clients already live within the major regions of Texas.
- They could do a comparison between their current school district and the one they are now strongly considering
- Moving can be a great hassle, so the client simply wants some data to convince them it may be worth it!
- For the example on the right, I would say the client is justified in strongly considering the move from Round Rock ISD (Austin) to Eanes ISD (Austin).

School District the Client Already Resides In:

What specific district are you interested in? round rock isd

*Predicted Average Graduation % for Classes of 2011 - 2017 + Rankings *

DistName	Graduated 4-Year (%)	Ranking(All Regions)	Ranking(Austin)
ROUND ROCK ISD	49.47	24	5

*Predicted Graduation % for Class of 2017 + Rankings *

DistName	Pred. 2021 Graduated 4-Year (%)	Ranking(All Regions)	Ranking(Austin)
ROUND ROCK ISD	48.205	29	7

School District the Client is Considering:

What specific district are you interested in? eanes isd

*Predicted Average Graduation % for Classes of 2011 - 2017 + Rankings *

DistName	Graduated 4-Year (%)	Ranking(All Regions)	Ranking(Austin)
EANES ISD	56.716429	8	2

*Predicted Graduation % for Class of 2017 + Rankings *

DistName	Pred. 2021 Graduated 4-Year (%)	Ranking(All Regions)	Ranking(Austin)
EANES ISD	58.21	4	1

Recommendations to the Clients (continued)

Let's say there's a new school district (population growth is very apparent in the major regions of Texas) that's been in operation for less than four years. If we have the school district's test results/features from its first graduating class, we can use Model 3 to predict the percentage of those students who will earn a college degree within four years after enrolling into a Texas college. Based on the results, clients may consider a move to the new district.

What high school class year is this prediction for?: 2019

Whats the district's Average ACT score?: 23.8

Whats the district's average ACT Participation (%): 55

Whats the district's average AP Passing (%): 76

Whats the district's average AP Exams Per Student?: 2.8

How many students enrolled into college?: 600

Whats the district's average SAT score?: 1175

Whats the district's average SAT Participation (%): 83

Whats the district's Wealth/ADA (\$): 950000



Predicted 2023 Graduation (%): 52.94