Variations in Trial Rate in NC

By Xinyuan Cao, Yifan Lu, Nicolson Panos, Jiayi Zhou

Introduction

Project Background

- Jury trials in criminal courts in the U.S. have suffered through blight for many years.
- The jury trial rate is an important component of a healthy criminal legal system.
  - Keep prosecutors accountable to the law and community values
  - Promote more awareness of the criminal courts and acceptance of their legitimacy
- If jury trials are too few, then a court system faces the danger of being considered illegitimate.

Project Goal

- Goal 1: Understand and describe the variations in criminal trial rates among counties in North Carolina over ten years with maps
- Goal 2: Explain the variations in criminal trial rates among counties in North Carolina with models

Data

- Staring data:
  - North Carolina felony case activity reports from 2010 to 2022
  - North Carolina criminal infraction case activity reports by the prosecutorial district and county from 2010 to 2022

Goal One: Calculate and Map the Variations

Part 1: Calculate the Variables

- Data: North Carolina felony case activity reports
  - Focusing on felony cases in each county
- Two variables of interest
  - Y1–Trial Conviction Rate: the percentage of trials as a percentage of all convictions obtained
  - Y2–Trial Disposition Rate: the percentage of trials as a percentage of total case disposal

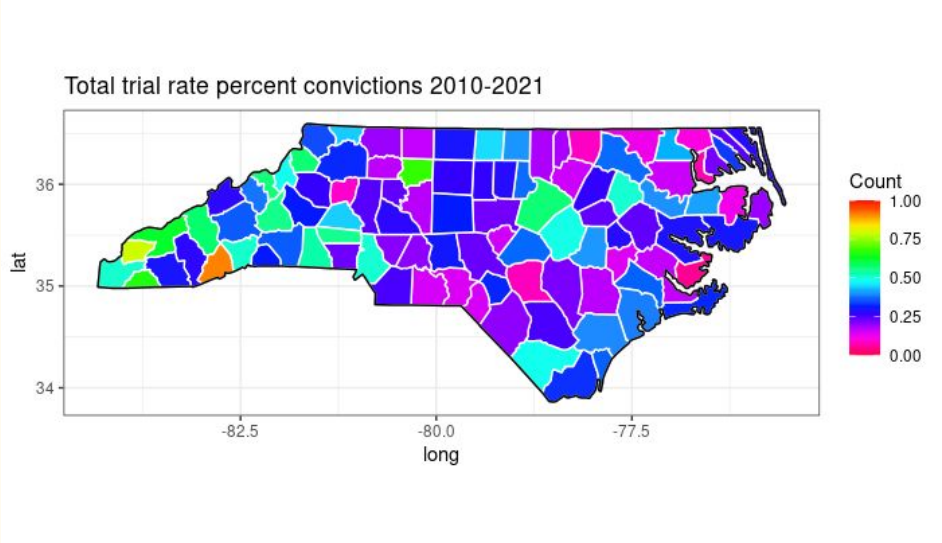
DC	DA	SC	AREA	FMT	CHARGI	CHARGE_TEXT	CRS_CASES_FILED	CRS_CASES_PENDING
NA	NA	NA	STATEWIDE	F	0910	MURDER OF AN UN	1	4
NA	NA	NA	STATEWIDE	F	0920	VOLUNTARY MANS	10	9
NA	NA	NA	STATEWIDE	F	0922	INVOLUNTARY MAN	62	57
NA	NA	NA	STATEWIDE	F	0930	MURDER	175	395
NA	NA	NA	STATEWIDE	F	0935	FIRST DEGREE MUR	491	968
NA	NA	NA	STATEWIDE	F	0940	SECOND DEGREE M	1	5
NA	NA	NA	STATEWIDE	F	0942	SECOND DEG MURI	19	20
NA	NA	NA	STATEWIDE	F	0943	SECOND DEG MURI	10	9
NA	NA	NA	STATEWIDE	F	0944	SECOND DEGREE M	85	100

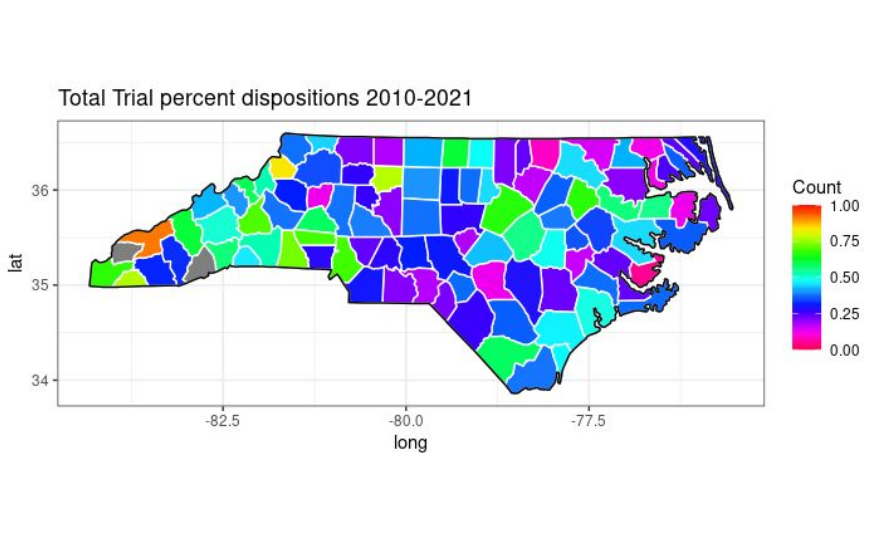
Part 2: Map - Data Preparation

- For the two variables of interest, organize them by aggregating all felonies per year per county
  - Year
  - County
  - Count

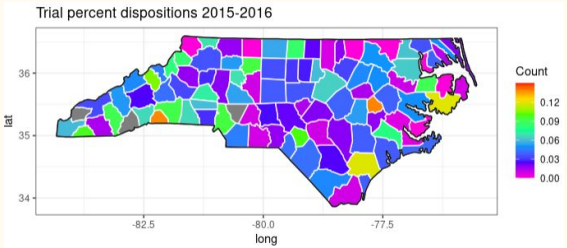
	Year	County	% Count
1	1011	ALAMANCE	0.0410918822
2	1112	ALAMANCE	0.0247958936
3	1213	ALAMANCE	0.0408910341
4	1314	ALAMANCE	0.0179074903
5	1415	ALAMANCE	0.0111043414
6	1516	ALAMANCE	0.0289129895
7	1617	ALAMANCE	0.0155134367
8	1718	ALAMANCE	0.0400976633
9	1920	ALAMANCE	0.0197697161
10	2021	ALAMANCE	0.0052580815
11	2122	ALAMANCE	0.0187994515
12	1011	ALEXANDER	0.0095351272
13	1112	ALEXANDER	0.0029827185

Part 2: Map - Grand Total

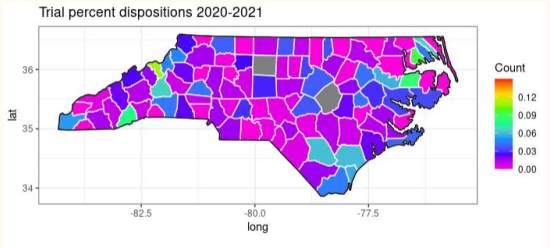




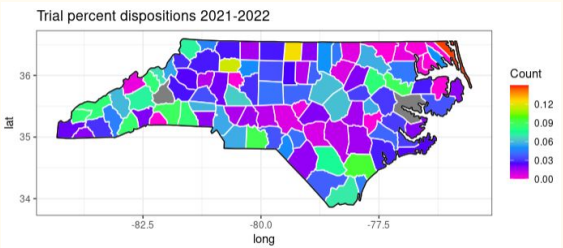
Part 2: Map - Trial Percent Disposition



Pre Covid

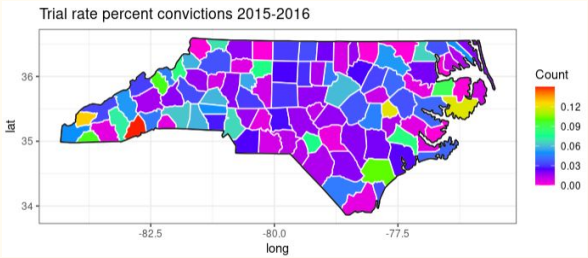


Covid

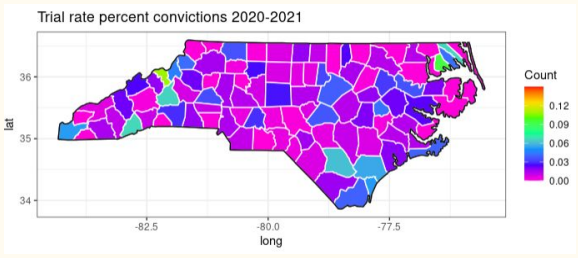


“Post” Covid

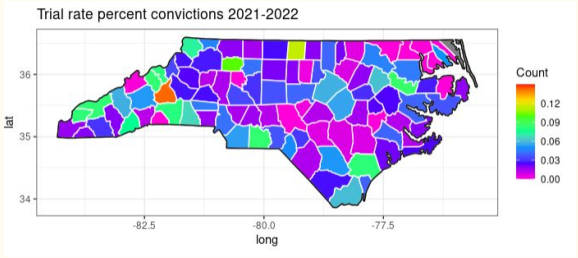
Part 2: Map - Trial Rate Percent Convictions



Pre Covid



Covid



“Post” Covid

Goal Two: Explain the Variations with Models

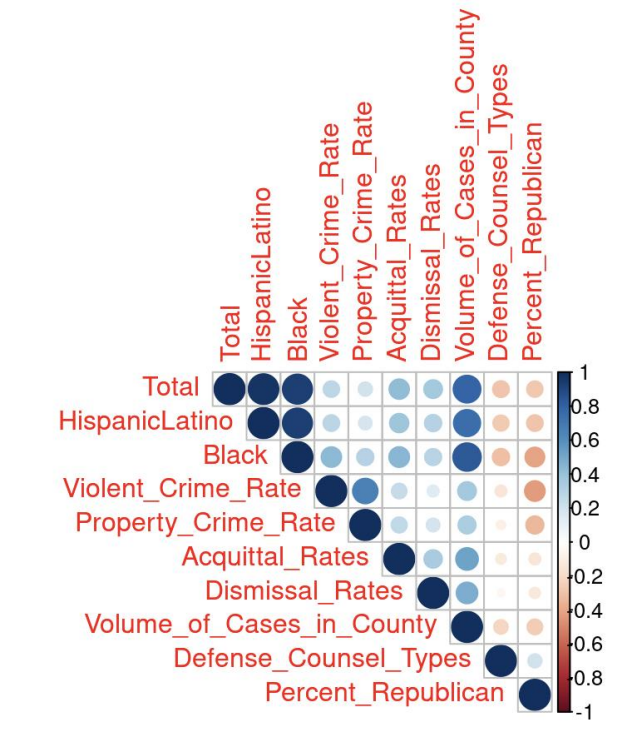
Collecting the Explanatory Variables

- Xinyuan
  - **Acquittal Rates** (percentage of jury trials that result in acquittal) , **Dismissal Rates** (percentage of total dispositions that happen through dismissals) , **Prosecutorial District**, **Judicial District**, **Volume of Cases in the County**, **Defense Counsel Type**.
- Jiayi
  - Crime rate: **Violent Crime Rate** and **Property Crime Rate** from the NC State Bureau of Investigation (Link to Database).
  - The racial and ethnic makeup of the county: **Total County Population**, **Percent of the Population that is Black**, and **Percent of the Population that is Hispanic/Latino** from United States Census Bureau (Link to Database).

Collecting the Explanatory Variables

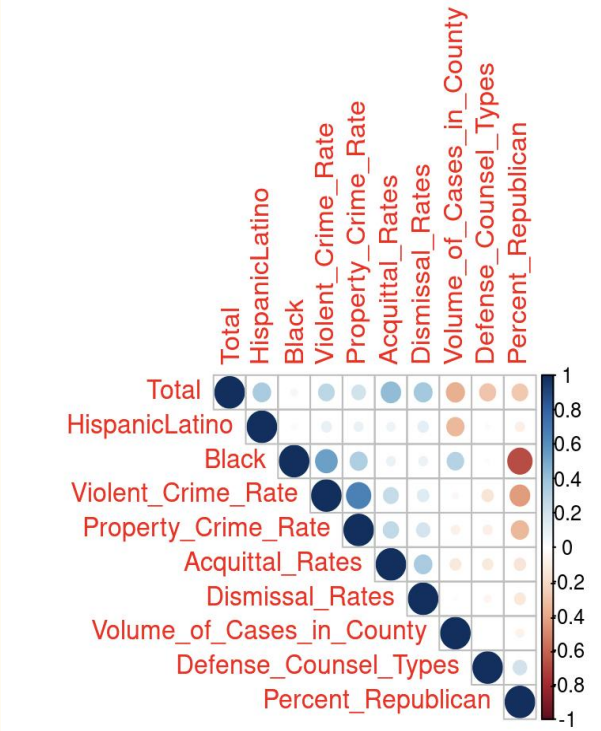
- Nico
  - **Percentage of Republican**: Going through Data of Voting Percentage Republic for Presidential Elections of 2008, 2012, 2016, and 2020 based on North Carolina County. Found from Election Data from the North Carolina State Board of Elections Database (Link to Database).

Data Exploration and Processing



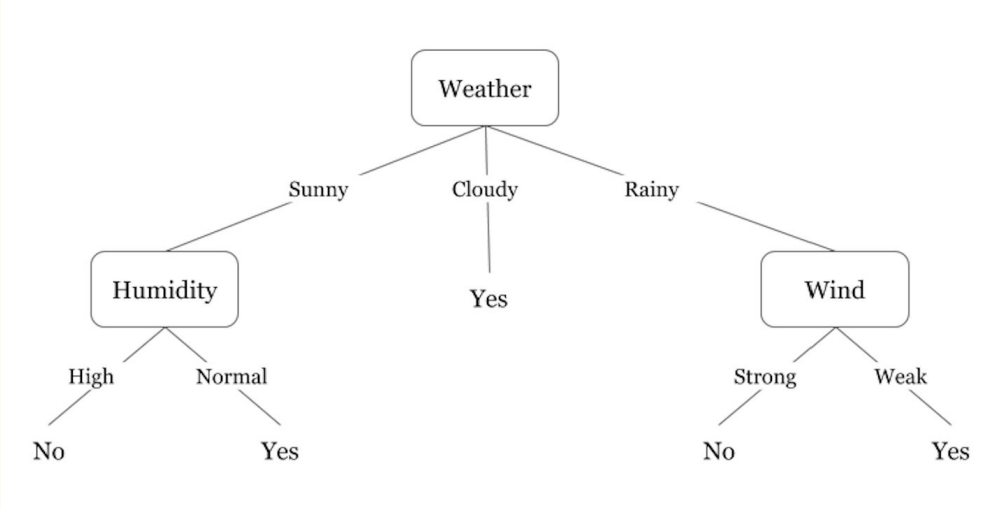
Changes of Explanatory Variables

- HispanicLatino=HispL/Total
- Black = Black/Total
- VolCases=Volcases/Total
- Use County to remove Prosecutorial District

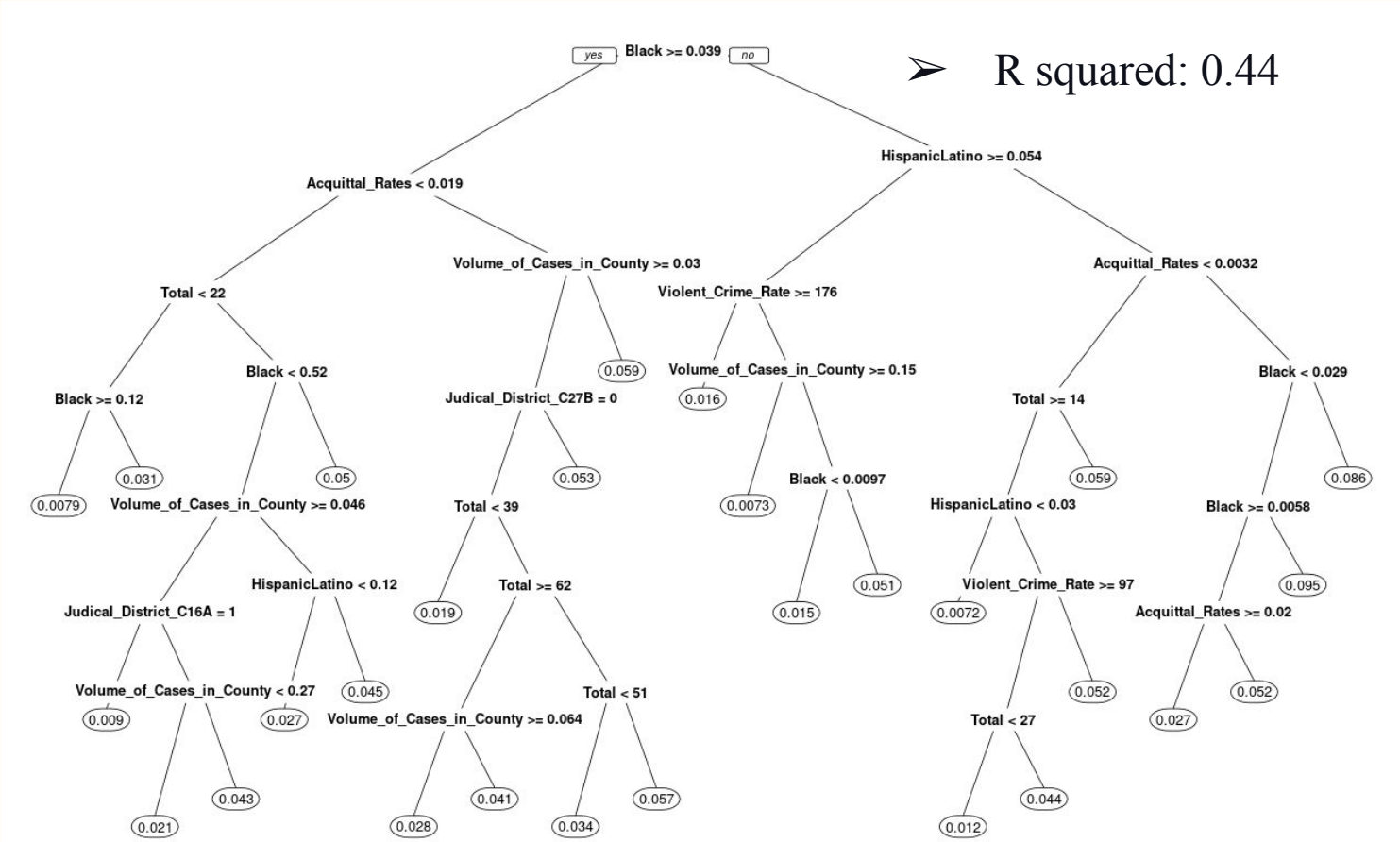


Model 1: Decision Tree

- A Decision Tree is a tree-like model of decisions and potential outcomes by splitting the dataset based on criteria.
- Trees are built by recursively evaluating different explanatory variables and using the one at each node that best splits the data and maximizes the R squared.
- To read a tree, we start from the root node and go to the next node based on the options provided by the edges.



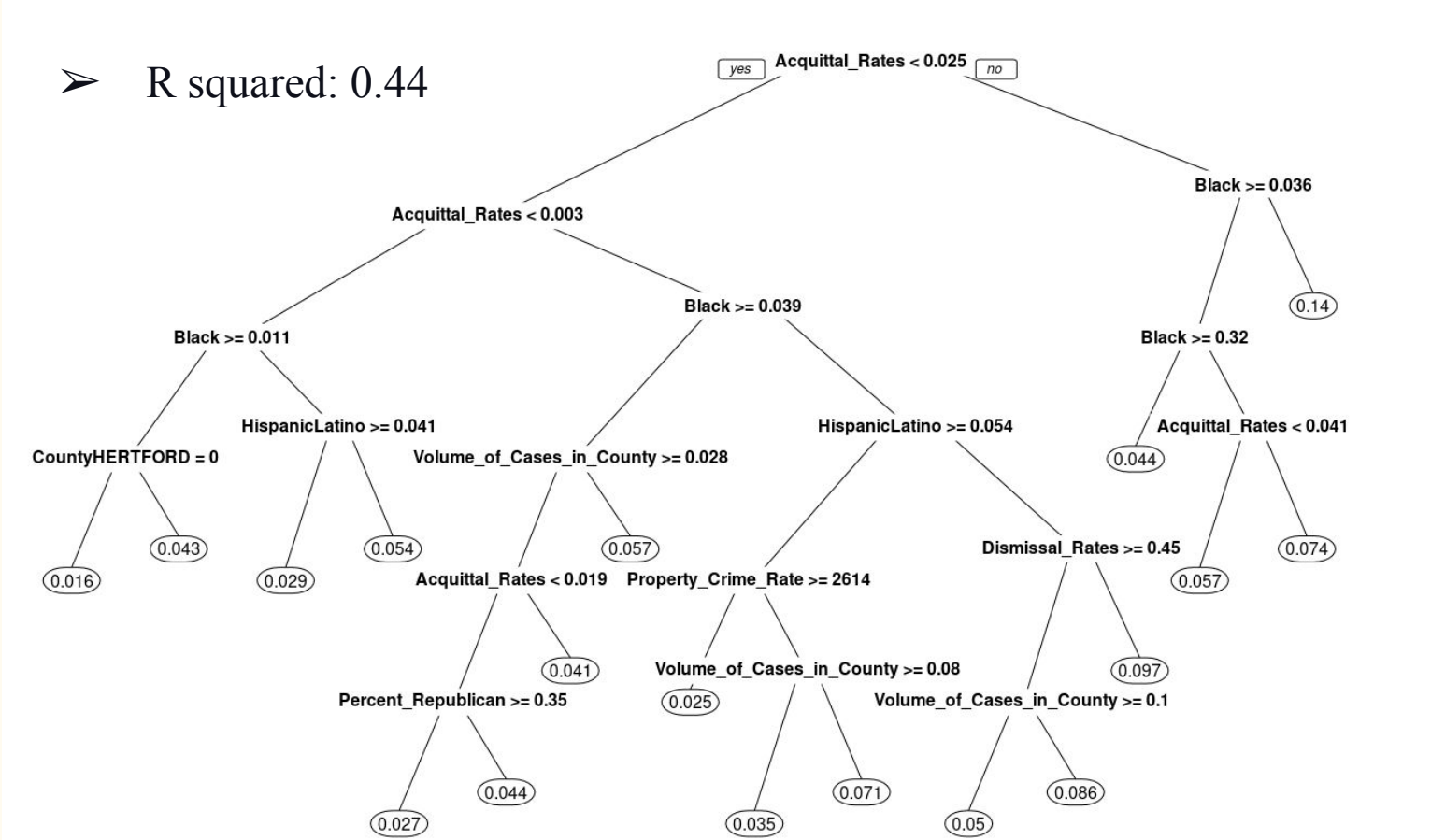
Model 1: Decision Tree for Trial Conviction Rate



- R squared: 0.44

Model 1: Decision Tree for Trial Disposition Rate

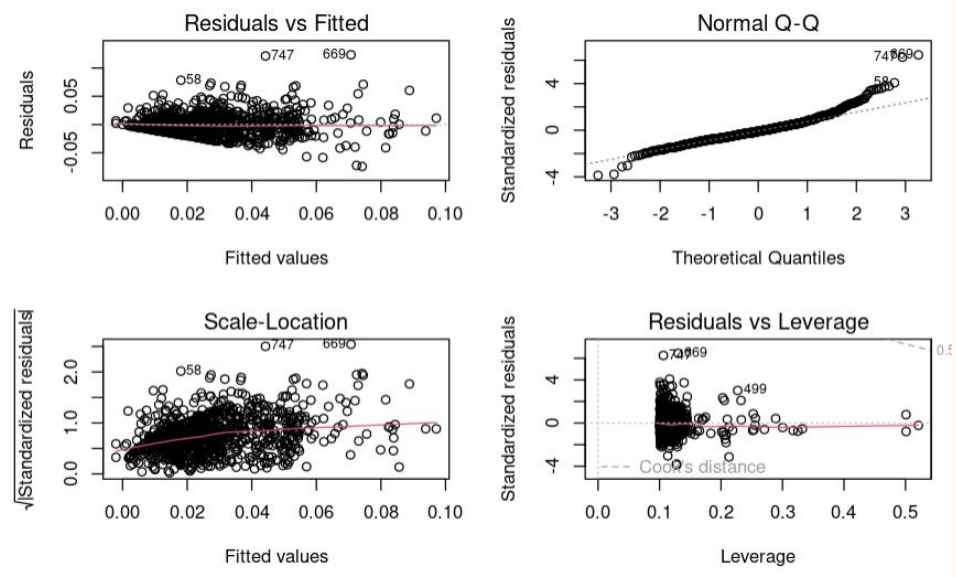
- R squared: 0.44



Model 2: Linear Models

- For Y1: The variables that proved to be of a significant level of interest were 3 of the indicator variables that represent the 100 counties in North Carolina, percent of the population in a county that is Hispanic/Latino, and the Acquittal Rates of the county.
- Multiple R-Squared Value: 0.4073

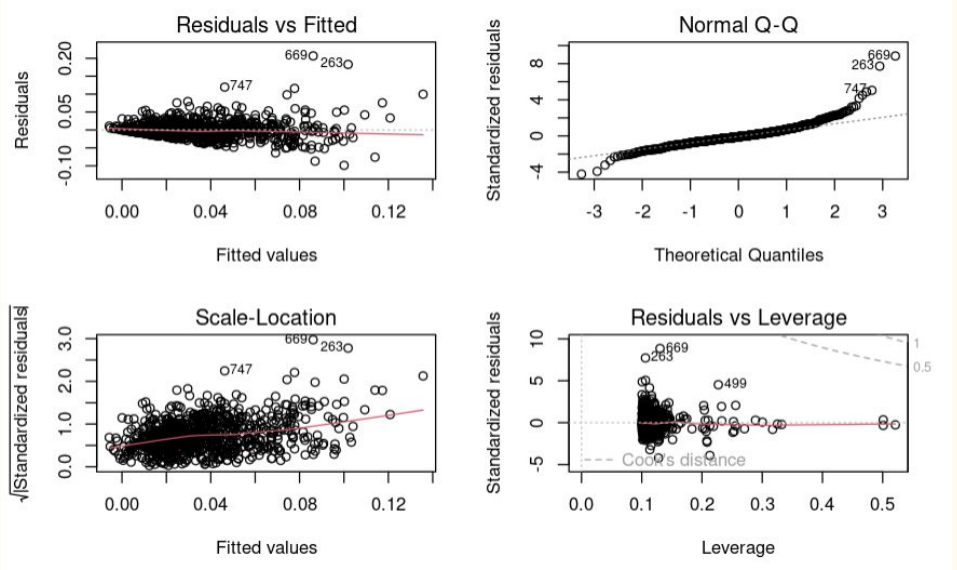
Checking Assumptions for Linear Model:



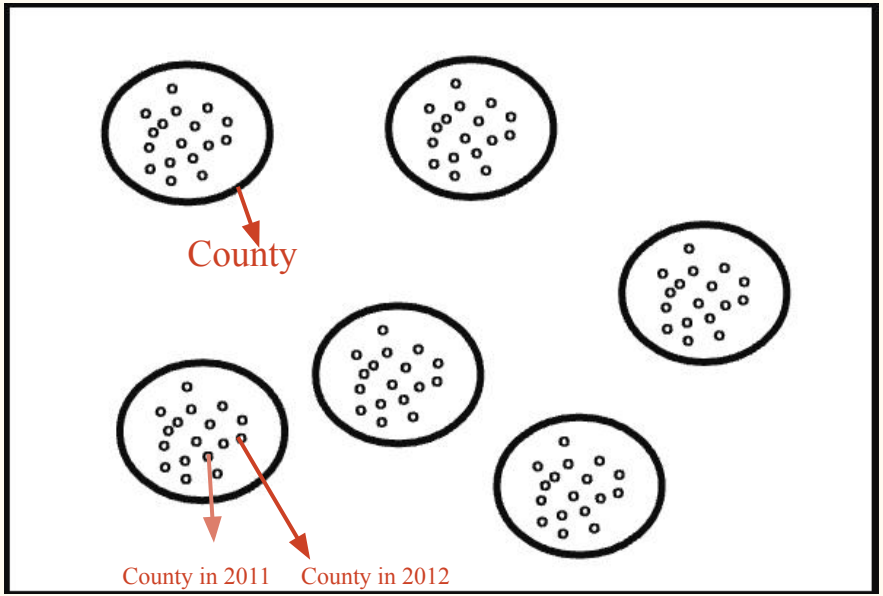
Model 2: Linear Models

- For Y2: The variables that proved to be of a significant level of interest were, again, 3 of the indicator variables that represent the 100 counties in North Carolina, the Acquittal Rates of the county, and the Hispanic/Latino percent of the population in a county.
- Multiple R-Squared Value: 0.4732

Checking Assumptions for Linear Model:



Model 3: Linear Mixed Effect Model



https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/

Linear Mixed Effects Model (LME) is an extension of simple linear models to allow both fixed and random Effects. (UCLA 2021)

Particularly used when there is non independence in the data, such as arises from a hierarchical structure. (UCLA 2021)

In our example, we consider County as a random effect, all the other variables are fixed effects.

LME model would let us look at the influence of our fixed effects within each County.

Model 3: Mixed Effect Model

Use County as a random effect in Y1(percentage of convictions)

Fixed Effect	Value	Std.Error	t-value	p-value
(Intercept)	0.7582200	0.7284828	1.040821	0.2983
Year	-0.0003597	0.0003619	-0.993841	0.3206
Total	-0.0000057	0.0000481	-0.118868	0.9054
HispanicLatino	-0.0829742	0.0737735	-1.124715	0.2610
Black	-0.0364689	0.0226683	-1.608807	0.1081
Violent_Crime_Rate	0.0000014	0.0000102	0.138194	0.8901
Property_Crime_Rate	0.0000006	0.0000016	0.358155	0.7203
Acquittal_Rates	0.2752450	0.0698234	3.942018	0.0001
Dismissal_Rates	-0.0001148	0.0043675	-0.026287	0.9790
Volume_of_Cases_in_County	-0.0208332	0.0139562	-1.492755	0.1359
Judicial District 21	0.0484111	0.0218429	2.216330	0.0317
Judicial District 29B	0.0334610	0.0101764	3.288110	0.0019
Judicial District 30A	0.0216647	0.0109580	1.977056	0.0540
Defense_Counsel_Types	0.0078377	0.0083800	0.935292	0.3545
Percent_Republican	-0.0101234	0.0123668	-0.818595	0.4133

Random Effects		
Formula: ~1 | County		
	Intercept	Residual
StdDev	0.01222225	0.02062361

- We expect that the average percentage of all convictions obtained by trial of each county may vary from the global average percentage by ± 0.0125.
- Within any given county, we expect the percentage of all convictions obtained by trial of each year may vary from the county's average percentage by ±0.021.
- Intraclass Correlation: 0.2691 (0.0122^2/0.0125^2+0.0206^2) 26.91% of the variation in Y1(percentage of all convictions obtained by trial) could be explained by differences in average percentage from county to county

Model 3: Mixed Effect Model

Use County as a random effect in Y2

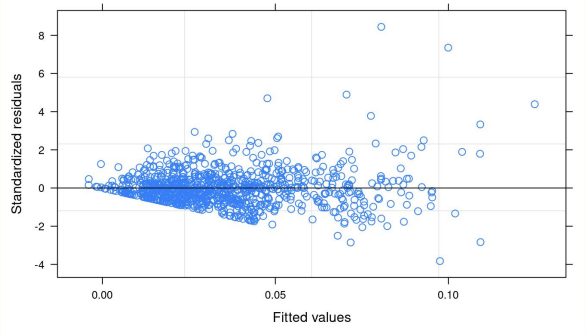
Fixed Effect	Value	Std.Error	t-value	p-value
(Intercept)	0.6665665	0.8709696	0.765315	0.4443
Year	-0.0003120	0.0004326	-0.721098	0.4711
Total	0.0000070	0.0000558	0.124769	0.9007
HispanicLatino	-0.1170092	0.0817728	-1.430905	0.1528
Black	-0.0406901	0.0252692	-1.610264	0.1077
Violent Crime Rate	0.0000080	0.0000123	0.650755	0.5154
Property Crime Rate	0.0000000	0.0000019	-0.009519	0.9924
Acquittal Rates	1.0322393	0.0850200	12.141129	0.0000
Dismissal Rates	-0.0062188	0.0053124	-1.170622	0.2421
Volume of Cases in County	-0.0250134	0.0168097	-1.488036	0.1371
Judicial District 29B	0.0399483	0.0111772	3.574071	0.0008
Judicial District 30A	0.0328102	0.0120222	2.729124	0.0090
Defense Counsel Types	0.0108927	0.0092003	1.183951	0.2425
Percent Republican	-0.0118074	0.0148643	-0.794347	0.4272

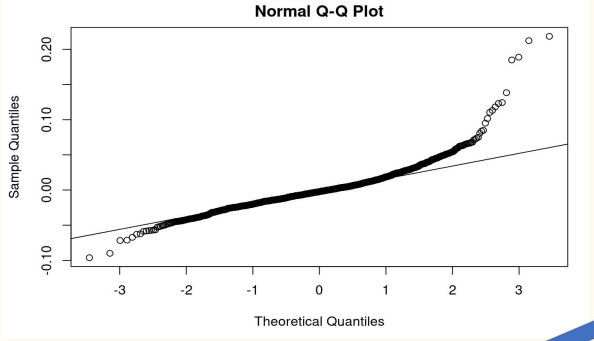
Random Effects		
Formula: ~1 | County		
	Intercept	Residual
StdDev	0.01183848	0.02514732

- We expect that the average percentage of all dispositions obtained by trial of each county may vary from the global average percentage by ± 0.01183.
- Within any given county, we expect the percentage of all dispositions obtained by trial of each year may vary from the county's average percentage by ±0.02515.
- Intraclass Correlation: 0.1812
  (0.011832^2/0.02515^2+0.01183^2)
  18.12% of the variation in Y2(percentage of all dispositions obtained by trial) could be explained by differences in average percentage from county to county

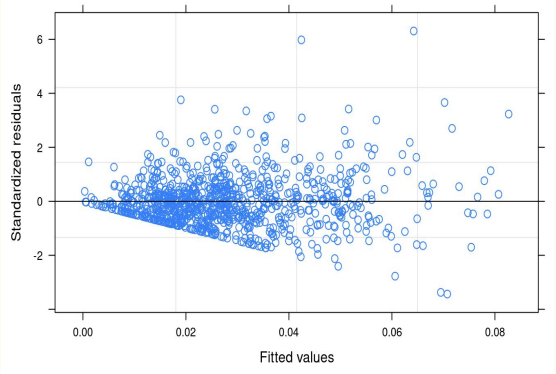
Model 3: Mixed Effect Model

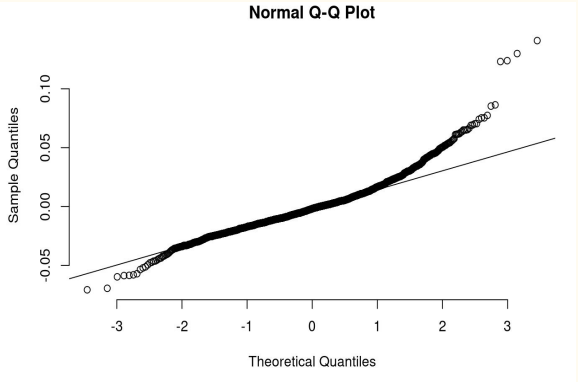
Check assumptions for Y1 model





Check assumptions for Y2 model





- Constant Variance not met
- Normality not met

Transform Y1 and Y2

Model 3: Mixed Effect Model

Use County as a random effect in sqrt(Y1)

Fixed Effect	Value	Std.Error	t-value	p-value
(Intercept)	2.8473002	2.3670903	1.202869	0.2294
Year	-0.0013439	0.0011761	-1.142632	0.2535
Total	0.0001195	0.0001581	0.755613	0.4501
HispanicLatino	-0.1901563	0.2467692	-0.770583	0.4412
Black	-0.1403891	0.0757193	-1.854072	0.0641
Violent Crime Rate	0.0000244	0.0000330	0.739444	0.4599
Property Crime Rate	0.0000039	0.0000051	0.765708	0.4441
Acquittal Rates	0.8240717	0.2252595	3.658322	0.0003
Dismissal Rates	0.0064379	0.0140957	0.456724	0.6480
Volume of Cases in County	-0.0813545	0.0451978	-1.799964	0.0722
Judicial District 29B	0.0926127	0.0341790	2.709637	0.0094
Defense Counsel Types	0.0204842	0.0281500	0.727680	0.4705
Percent Republican	-0.0346719	0.0400751	-0.865172	0.3872

Random Effects		
Formula: ~1 | County		
	Intercept	Residual
StdDev	0.03859531	0.06637771

- We expect that the average sqrt percentage of all convictions obtained by trial of each county may vary from the global average percentage by ± 0.038.
- Within any given county, we expect the sqrt percentage of all dispositions obtained by trial of each year may vary from the county's average percentage by ±0.0664.
- Intraclass Correlation: 0.2526
  (0.03859^2/0.03859^2+0.066378^2)
  25.26% of the variation in sqrt Y2(percentage of all dispositions obtained by trial) could be explained by differences in average sqrt percentage from county to county

Model 3: Mixed Effect Model

Use County as a random effect in sqrt(Y2)

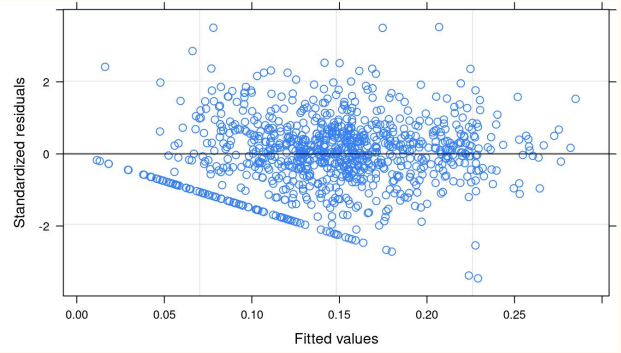
Fixed effect	Value	Std.Error	t-value	p-value
(Intercept)	2.3897554	2.3578679	1.013524	0.3111
Year	-0.0011124	0.0011715	-0.949567	0.3426
Total	0.0001684	0.0001560	1.079567	0.2807
HispanicLatino	-0.2371311	0.2393423	-0.990761	0.3221
Black	-0.1348166	0.0735328	-1.833421	0.0671
Violent Crime Rate	0.0000358	0.0000330	1.085749	0.2779
Property Crime Rate	0.0000019	0.0000051	0.366932	0.7138
Acquittal Rates	2.6881766	0.2258661	11.901641	0.0000
Dismissal Rates	-0.0105745	0.0141287	-0.748444	0.4544
Volume of Cases in County	-0.0932547	0.0451601	-2.064982	0.0392
Judicial District 29B	0.0948491	0.0330255	2.871993	0.0061
Judicial District 30A	0.0721553	0.0355635	2.028914	0.0483
Defense Counsel Types	0.0220710	0.0271960	0.811555	0.4212
Percent Republican	-0.0257058	0.0400192	-0.642337	0.5208

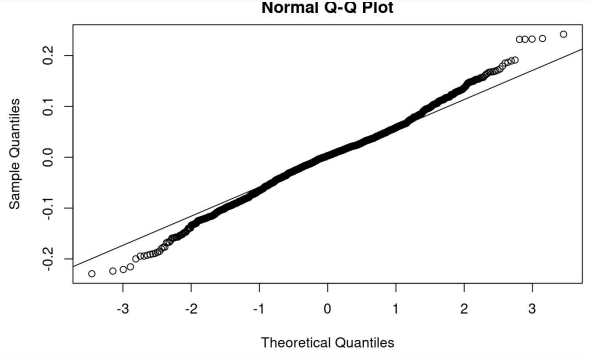
Random Effects		
Formula: ~1 | County		
	Intercept	Residual
StdDev	0.03676679	0.06661673

- We expect that the average sqrt percentage of all dispositions obtained by trial of each county may vary from the global average percentage by ± 0.0367.
- Within any given county, we expect the sqrt percentage of all dispositions obtained by trial of each year may vary from the county's average percentage by ±0.066.
- Intraclass Correlation: 0.2329
  (0.0367^2/0.0367^2+0.0666^2)
  23.29% of the variation in sqrt Y2(percentage of all dispositions obtained by trial) could be explained by differences in average sqrt percentage from county to county

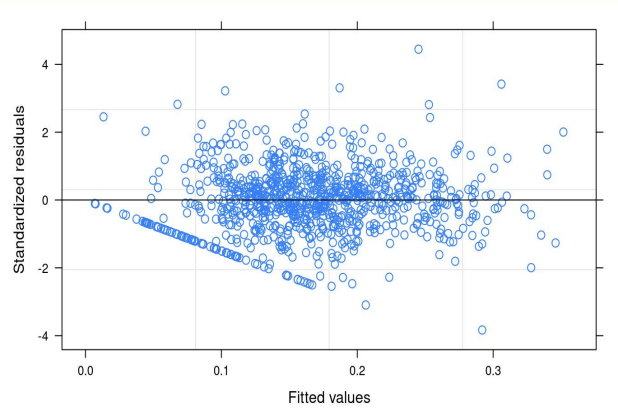
Model 3: Mixed Effect Model

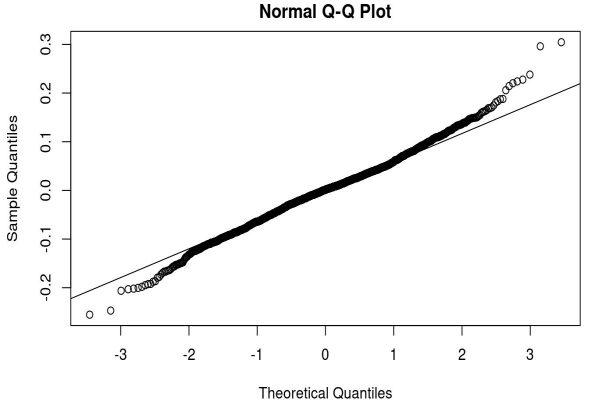
Check assumptions for Y1 transform model





Check assumptions for Y2 transform model





Constant Variance and Normality improves a lot

Conclusion

Conclusions

- From looking at the three models, the Mixed Effects Model proved the best for understanding the variation in the data and interpretation. We could fix the county instead of worrying about 100 different indicator variables that proved to have varying levels of significance.
- Based on the three models, we conclude that **Acquittal Rates**, **Judicial District**, **Percent of the population that is Black**, **Percent of the population that is Hispanic/Latino**, and **Volume of Cases in the County** are related to the variations in criminal trial rates among counties in North Carolina.
- Based on the Mixed Effects Model, **Acquittal Rates** and **Percent of the population that is Black** have the two highest values in coefficients to predict the variations in criminal trial rates among counties in North Carolina.