Effect of the 2021 Virginia Minimum Wage Hike on the Employment Level of Food Service Workers

By Jacob Langley

2023-01-16

Abstract

- On May 1st 2021, Virginia raised its minimum wage from the federal minimum of $7.25 to $9.50 and again to $11.00 on January 1st 2022
- In the same timeframe Virginia's neighbor, North Carolina, kept the federal minimum of $7.25
- This created a "natural experiment" to compare Virginia with its neighbor to see if the hike had an effect on the state's employment level

Abstract

- I use a Difference In Differences (DID) regression with North Carolina as a control, and logistic and lasso regressions for variable selection guidance
- The final model shows that neither minimum wage hike has had a statistically significant effect on the employment level among food and beverage workers in Virginia thus far

Research Context

- According to a report by The Congressional Budget Office in 2019 an increase to a $15 minimum wage would raise 1.3 million people out of poverty but with a two thirds chance employment would decrease by anywhere between 0 and 3.7 million workers
- 3.7 million would be about a 65% increase in the unemployment rate
- This potential employment loss statistic is so broad it conveys very little insight into whether or not raising the federal minimum wage is dangerous for the economy

Research Context

- It is vital that economics as a field is able to give an answer when someone asks to what degree minimum wage hikes effect employment
- The purpose of this project is to identify a specific example of a recent minimum wage change and give a more narrow response on its effect on the employment level

Research Context

2021 Economics Nobel Prize:

"Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania" by David Card and Alan Krueger in 1994

- Examined New Jersey's minimum wage hike in 1992 from \$4.25 to \$5.05 and its effect on the employment level among fast food workers in the area
- The paper used a DID method to compare the employment level on either side of the NJ/PA border before and after the change

Research Context

What is DID Regression?

- DID regression is a method that compares a control group and a subject group before and after a treatment is applied to the subject
- Compares differences over time and differences between the groups
- If there is a significant difference between the subject and control group after the treatment, it suggests the treatment had an impact
- The null hypothesis is that the treatment had no effect
- Utilizes cross-sectional data to mimic a traditional randomized control trials for “natural experiments”

Research Context

- Taking inspiration from Card and Krueger, I have also utilized a DID analysis for this project
- My observations are on the county level where Card and Krueger's observations were individual restaurants
- Like Card and Krueger, I used a neighbor state as the control
- DID as a method came under scrutiny in 2019 from a paper published in The Review of Economics and Statistics

Methodology

Initial Regression

- The initial regression includes all available demographic information in my created dataset
- The major question at this point is how good of a control North Carolina is for Virginia in each of these variables
- At this point we do not know which variables contain useful information and which are just measuring differences between the states

Methodology

Variable Selection

- Treatment (Minimum Wage Hike) is similar enough to control that if there are proxies taking significance away from Control, they would also be shifting significance away from the Minimum Wage Hike
- The primary focus during the variable selection is to identify which variables are good predictors of employment and not just proxies for capturing the inherent differences between the two states
- Those variables that do a good job at categorizing the two states and a poor job of predicting the employment level can likely be dropped from the model

Methodology

Variable Selection

- The secondary goal of the variable selection is to identify whether Percent White or Percent Black is more relevant to the model since they cannot both be included because of their high inverse correlation (-96%)

Methodology

Variable Selection: Logistic Regression

- Logistic regression is utilized to identify which variables are good at discriminating between the two states
- If a variable is very effective at distinguishing between the two states it may not be as useful of a predictor of employment in the final model

Methodology

Variable Selection: Lasso Regression

- Lasso regression adds a punishment term and forces weaker predictors of the dependent variable to zero
- As the power of the punishment term increases the less useful predictors get pushed to zero first
- This method is especially useful when working with small datasets because it is less sensitive to outliers
- Between the Logistic and Lasso Regression we should get a good idea of which variables should be kept in the model for a stronger case

Initial Regression

Table 4: Original OLS Estimates of the Effect of Minimum Wage on the Employment Level in VA

	Dependent variable:			
	Employment among Food and Beverage Retailers in VA and NC			
	2019-2021	2019-2021	2019-2022	2019-2022
	(1)	(2)	(3)	(4)
Time	103.932	105.244	100.688	103.335
	(212.090)	(212.868)	(208.852)	(209.625)
Control	−465.194**	−484.558**	−448.790**	−467.481**
	(211.483)	(212.001)	(209.056)	(209.615)
Minimum Wage Hike	−83.746	−85.009	−98.450	−101.498
	(285.939)	(286.984)	(279.417)	(280.453)
Average age (grouped)	−477.285****	−516.424****	−464.212****	−500.939****
	(131.521)	(129.825)	(126.973)	(125.537)
Gender	−1,833.666	−1,850.932	−937.176	−933.759
	(3,583.782)	(3,598.168)	(3,428.171)	(3,442.005)
Percent White	−13.036***		−12.956***	
	(4.683)		(4.510)	
Percent Black		10.878**		10.842**
		(4.825)		(4.645)
Percent Hisp/Latino	30.304*	30.580*	32.960*	33.356*
	(17.222)	(17.360)	(16.838)	(16.969)
Population Density	−0.240***	−0.233***	−0.248****	−0.243***
	(0.075)	(0.075)	(0.073)	(0.074)
College Graduation Rate	105.830****	106.196****	105.191****	105.573****
	(10.351)	(10.606)	(10.134)	(10.377)
Constant	5,190.012**	4,316.541*	4,601.541**	3,701.801*
	(2,156.160)	(2,212.769)	(2,049.545)	(2,099.113)
Observations	369	369	380	380
R^2	0.349	0.344	0.349	0.344
Adjusted R^2	0.332	0.328	0.333	0.328
Residual Std. Error	1,364.857 (df = 359)	1,369.843 (df = 359)	1,351.731 (df = 370)	1,356.771 (df = 370)
F Statistic	21.358**** (df = 9; 359)	20.913**** (df = 9; 359)	22.005**** (df = 9; 370)	21.537**** (df = 9; 370)
Note:				* p<0.1; ** p<0.05; *** p<0.01; ****p<0.001

Initial Regression

- The DID Treatment Variable "Minimum Wage Hike" was found to be not significant
- The Control variable was found to be significant at the .05 threshold implying some of the employment level can be explained through the difference between the states

Variable Selection: Logistic Regression

Table 5: Logistic Regression to Evaluate the Classification Power of the Variables

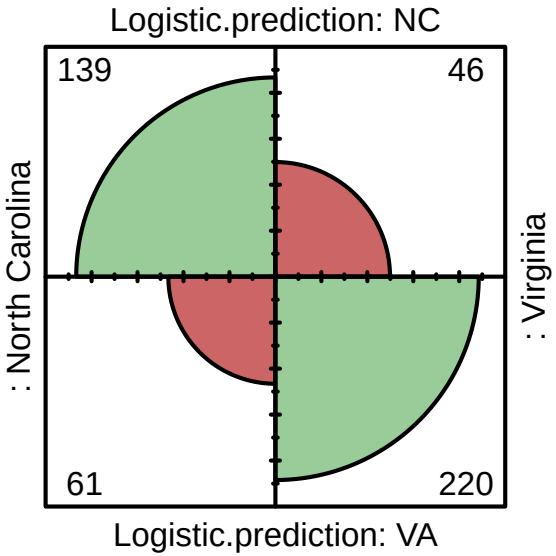
	Dependent variable:			
	Control: North Carolina = 0, Virginia = 1			
	2019-2021	2019-2021	2019-2022	2019-2022
	(1)	(2)	(3)	(4)
Average age (grouped)	−0.024	0.024	−0.024	0.024
	(0.222)	(0.219)	(0.222)	(0.219)
Gender	−30.221****	−30.889****	−30.221****	−30.889****
	(7.025)	(7.079)	(7.025)	(7.079)
Percent White	0.012*		0.012*	
	(0.007)		(0.007)	
Percent Black		−0.008		−0.008
		(0.008)		(0.008)
Percent Hisp/Latino	−0.341****	−0.340****	−0.341****	−0.340****
	(0.043)	(0.043)	(0.043)	(0.043)
Population Density	0.003****	0.003****	0.003****	0.003****
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
College Graduation Rate	0.037**	0.039**	0.037**	0.039**
	(0.018)	(0.019)	(0.018)	(0.019)
Constant	15.383****	16.376****	15.383****	16.376****
	(3.941)	(3.964)	(3.941)	(3.964)
Observations	466	466	466	466
Log Likelihood	−229.935	−230.847	−229.935	−230.847
Akaike Inf. Crit.	473.870	475.695	473.870	475.695
Note:	* p<0.1; ** p<0.05; *** p<0.01; ****p<0.001			

Variable Selection

Logistic Regression

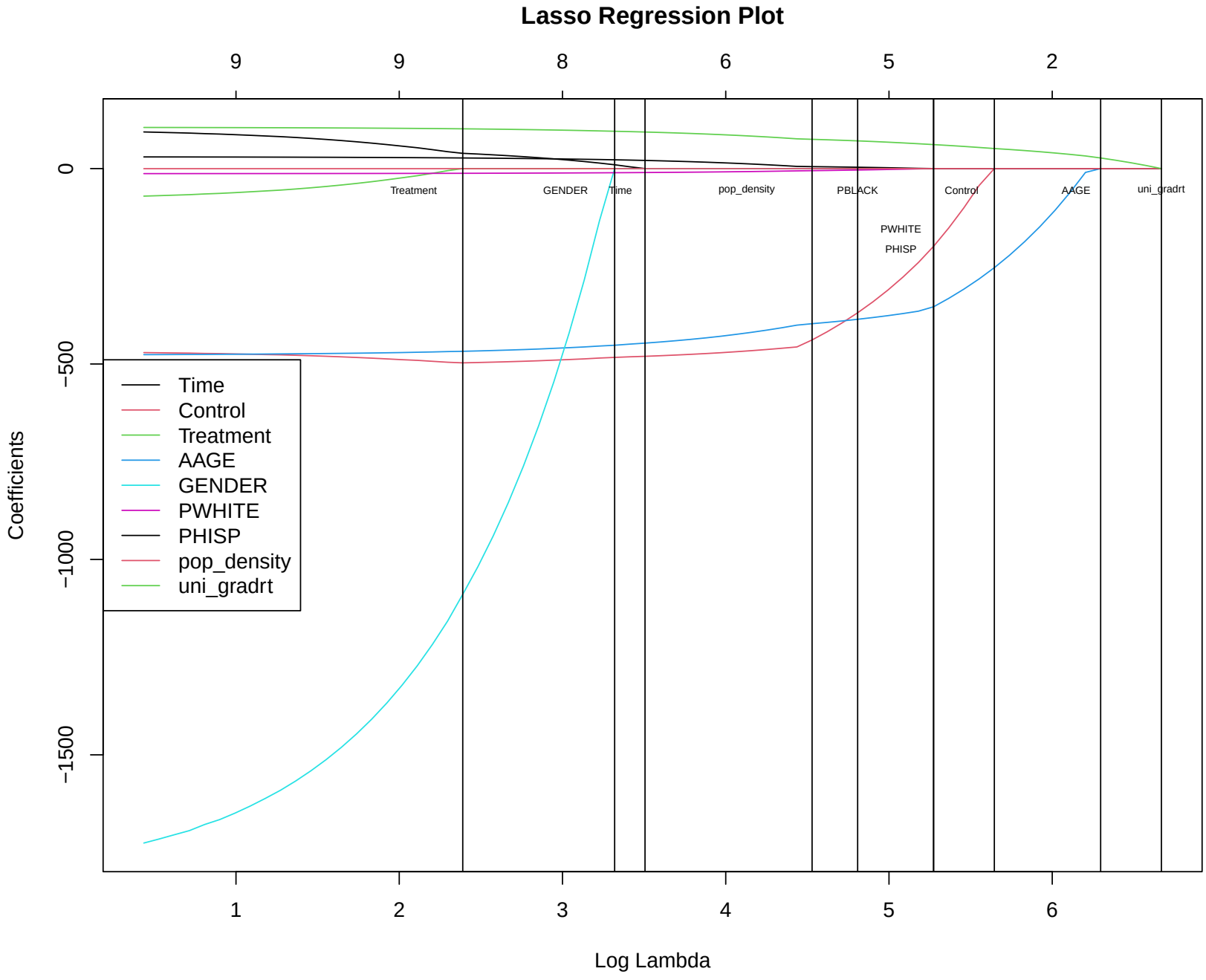
- Table 5 shows that average age, percent black, and percent white were not significant to the categorization
- College Graduation Rate was significant at the p$<$.05 level and the rest were highly significant
- The Logistic regression had a 23% error rate in its prediction for all four versions.

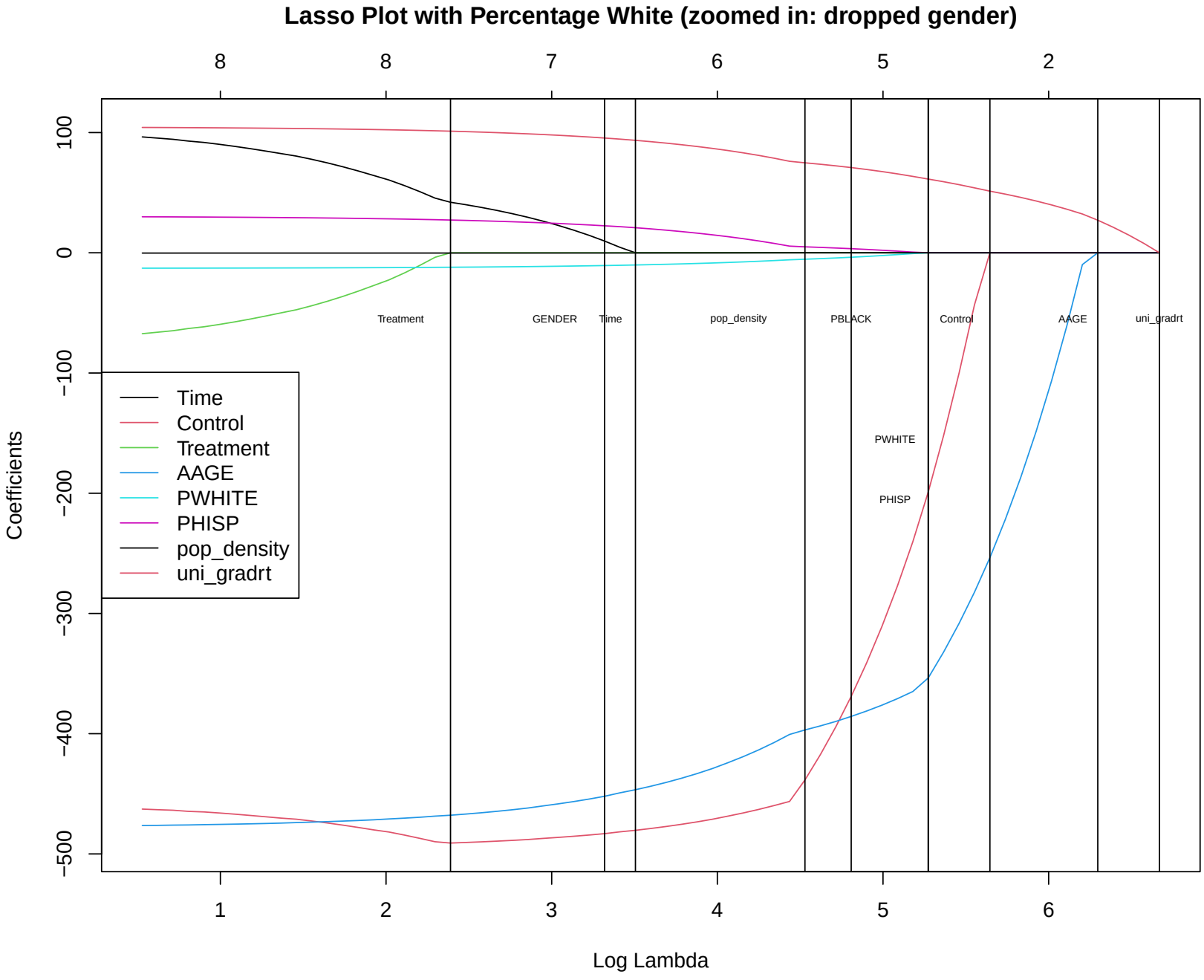
Confusion Matrix

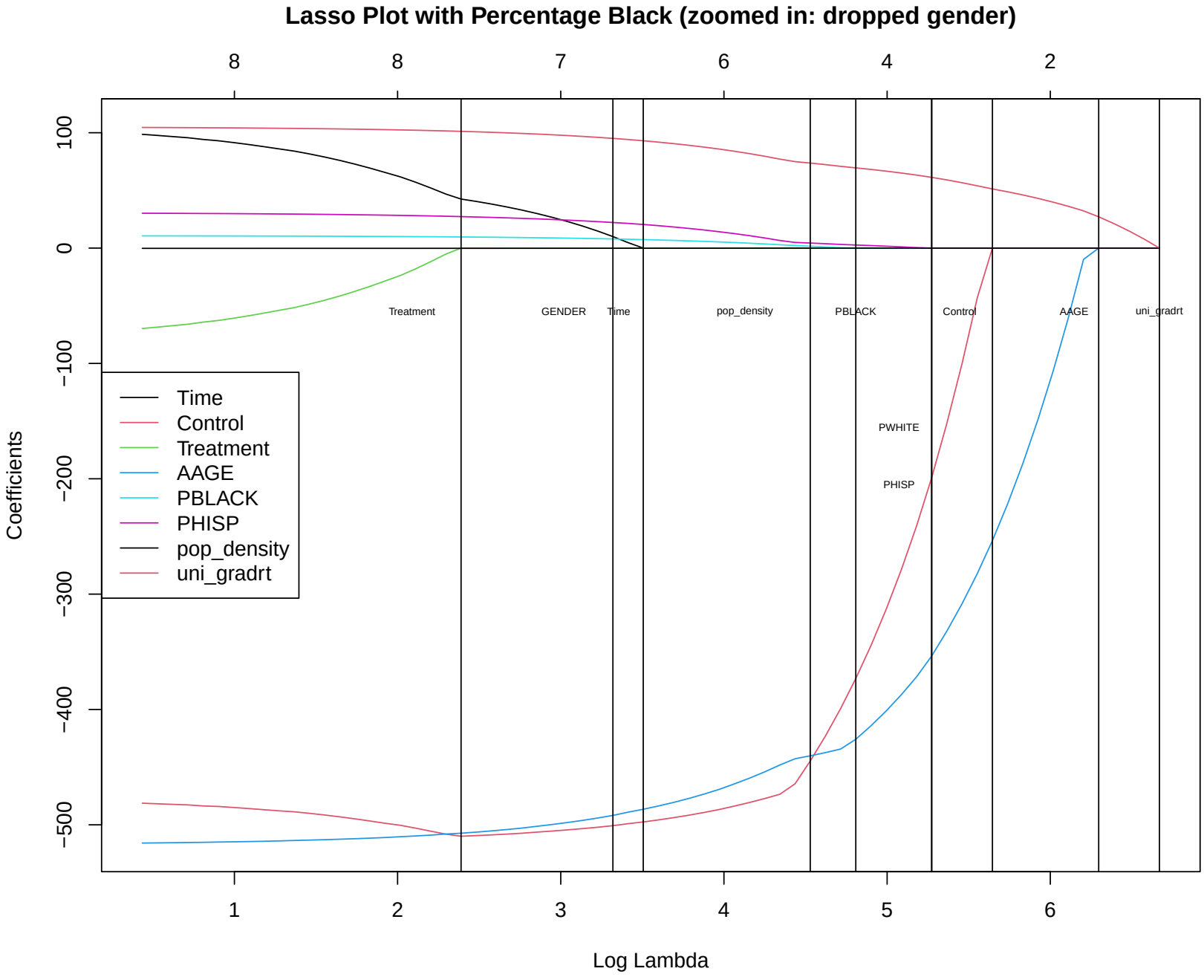


Variable Selection: Lasso Regression

	log(lambda)
Treatment	2.39
Gender	3.32
Time	3.51
Pop_Density	4.53
Perc_Black	4.81
Perc_White	5.27
Perc_Hisp	5.27
Control	5.65
Ave_Age	6.3
Uni_Grad_Rt	6.67


Lasso Regression Plot
9
9
8
6
5
2
0
−500
−1000
−1500
Coefficients
Treatment
GENDER
Time
pop_density
PBLACK
Control
AAGE
uni_gradrt
PWHITE
PHISP
Time
Control
Treatment
AAGE
GENDER
PWHITE
PHISP
pop_density
uni_gradrt
1
2
3
4
5
6
Log Lambda


Lasso Plot with Percentage White (zoomed in: dropped gender)
8
8
7
6
5
2
100
0
−100
−200
−300
−400
−500
Coefficients
Treatment
GENDER
Time
pop_density
PBLACK
Control
AAGE
uni_gradrt
PWHITE
PHISP
Time
Control
Treatment
AAGE
PWHITE
PHISP
pop_density
uni_gradrt
1
2
3
4
5
6
Log Lambda


Lasso Plot with Percentage Black (zoomed in: dropped gender)
8
8
7
6
4
2
100
0
−100
−200
−300
−400
−500
Coefficients
Treatment
GENDER
Time
pop_density
PBLACK
Control
AAGE
uni_gradrt
PWHITE
PHISP
Time
Control
Treatment
AAGE
PBLACK
PHISP
pop_density
uni_gradrt
1
2
3
4
5
6
Log Lambda

Variable Selection

Lasso Regression

- These results suggest that Percentage Black is a weaker predictor than Percentage White of the employment level among food and beverage retailers
- This creates a natural lambda selection at log(lambda) $= 5.27$ meaning we would drop every variable left of the double bar
- This solves our multicolinearity violation between Percent White and Percent Black
- Out of the other four variables to the left of the selected lambda; Treatment, Gender, Time, and Population Density, only Population Density was significant in the original model

Variable Selection

Final Model

- The DID variables will need to stay to maintain the assumptions of DID regression
- Variable Selection has allowed us to drop
  - Gender
  - Percent Hisp/Latino
  - Population Density
  - Percent Black

Results

Table 8: Final Model

	Dependent variable:		
	Employment among Food and Beverage Retailers		
	Original	Final Model	Final with 2022 employment
	(1)	(2)	(3)
Time	103.932	115.606	106.746
	(212.090)	(214.208)	(211.138)
Control	−465.194**	−631.897***	−630.617***
	(211.483)	(203.816)	(202.076)
Minimum Wage Hike	−83.746	−87.194	−100.218
	(285.939)	(288.896)	(282.709)
Average age (grouped)	−477.285****	−471.180****	−456.023****
	(131.521)	(115.022)	(111.188)
Gender	−1,833.666		
	(3,583.782)		
Percent White	−13.036***	−10.379**	−10.397**
	(4.683)	(4.662)	(4.502)
Percent Hisp/Latino	30.304*		
	(17.222)		
Population Density	−0.240***		
	(0.075)		
College Graduation Rate	105.830****	87.974****	87.734****
	(10.351)	(8.528)	(8.385)
Constant	5,190.012**	4,500.775****	4,371.644****
	(2,156.160)	(992.585)	(970.073)
Observations	369	369	380
R^2	0.349	0.329	0.327
Adjusted R^2	0.332	0.318	0.316
Residual Std. Error	1,364.857 (df = 359)	1,379.735 (df = 362)	1,368.186 (df = 373)
F Statistic	21.358**** (df = 9; 359)	29.567**** (df = 6; 362)	30.245**** (df = 6; 373)
Note:			* p<0.1; ** p<0.05; *** p<0.01; ****p<0.001

Results

The variables that were weak predictors of employment, or that were acting as proxies for Control have been dropped; leaving us with a more predictive model that only holds the necessary variables shown in Table 8. In this final model, the Minimum Wage Hike is still shown to be not significant, meaning it has not effected the employment level in the state. Updating the employment data to 2022 shows the same result.

Results

The remaining significant variables can be interpreted as follows:

- Grouped Average Age: for every 5 years the average age of the population is older, the employment level among food and beverage retailers decreases by 471 positions, working out to be 94 positions per average year older.
- Percent White: for each percentage point higher the employment among food and beverage retailers level decreases by 10 positions
- College Graduation Rate: for every percentage point higher of people with a college degree, food and beverage retailer employment in the county increases by 88 jobs.
- State difference: employment among food and beverage retailers in VA lowers by 664 when compared to NC.

Discussion

- Selection bias problem
- Longer lag effect
- Virginia raised the minimum wage again to $12 on January 1 of this year and are planning two more: $13.50 in 2025 and $15 in 2026
- A future study could collect the missing data manually and do another analysis since it appears there will be ample opportunity

Sources

Congressional Budget Office , & Alsalam, N., The Effects on Employment and Family Income of Increasing the Federal Minimum Wage (n.d.).

Difference-in-difference estimation. publichealth.columbia.edu. (n.d.). Retrieved January 16, 2023, from https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation

Doosa, G. (2021, June 14). The mathematical background of lasso and Ridge regression. Medium. Retrieved January 16, 2023, from https://medium.com/codex/mathematical-background-of-lasso-and-ridge-regression-23b74737c817#:~:text=Lasso%20or%20L1%20Regression%3A,directly%20performs%20f

Ferman, B., & Pinto, C. (2019). Inference in differences-in-differences with few treated groups and heteroskedasticity. The Review of Economics and Statistics, 101(3), 452–467. https://doi.org/10.1162/rest_a_00759

Sources

U.S. Bureau of Labor Statistics. (n.d.). QCEW data files. U.S. Bureau of Labor Statistics. Retrieved December 7, 2022, from https://www.bls.gov/cew/downloadable-data-files.htm

U.S. Census Bureau. (n.d.). Educational Attainment. Explore census data. Retrieved December 7, 2022, from https://data.census.gov/table?t=Educational%2BAttainment&g=0400000US37%240500000&y=2020&tid=ACSST5Y2020.S1501

US Census Bureau. (2021, October 8). County population by characteristics: 2010-2019. Census.gov. Retrieved December 7, 2022, from https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-detail.html

US Census Bureau. (2022, June 30). County population by characteristics: 2020-2021. Census.gov. Retrieved December 7, 2022, from https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html

Sources

Virginia Open Data Portal. Tyler Data & Insights. (n.d.). Retrieved December 7, 2022, from https://data.virginia.gov/

https://www.bls.gov/news.release/empsit.t01.htm

Thank You

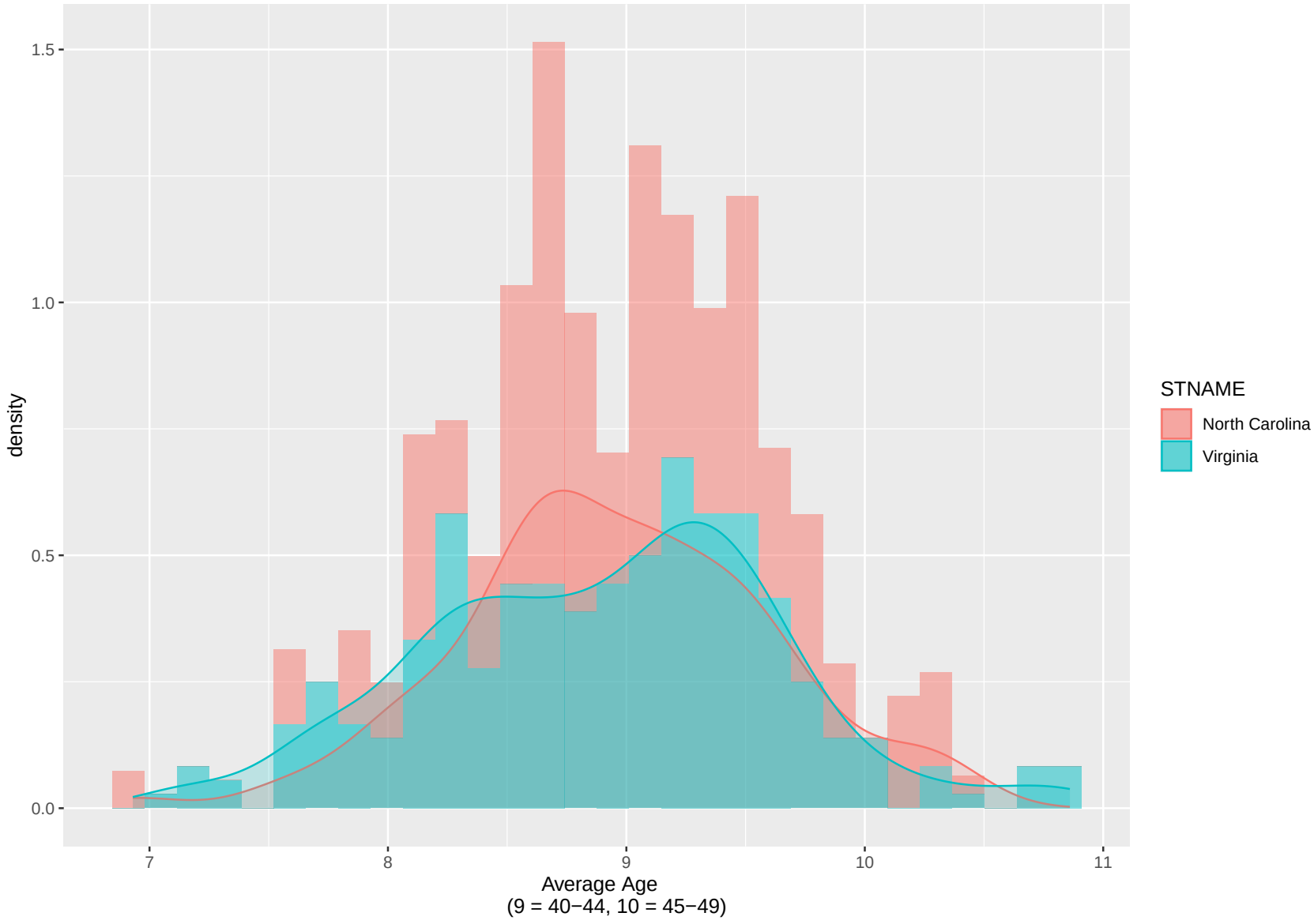
Questions

Appendix

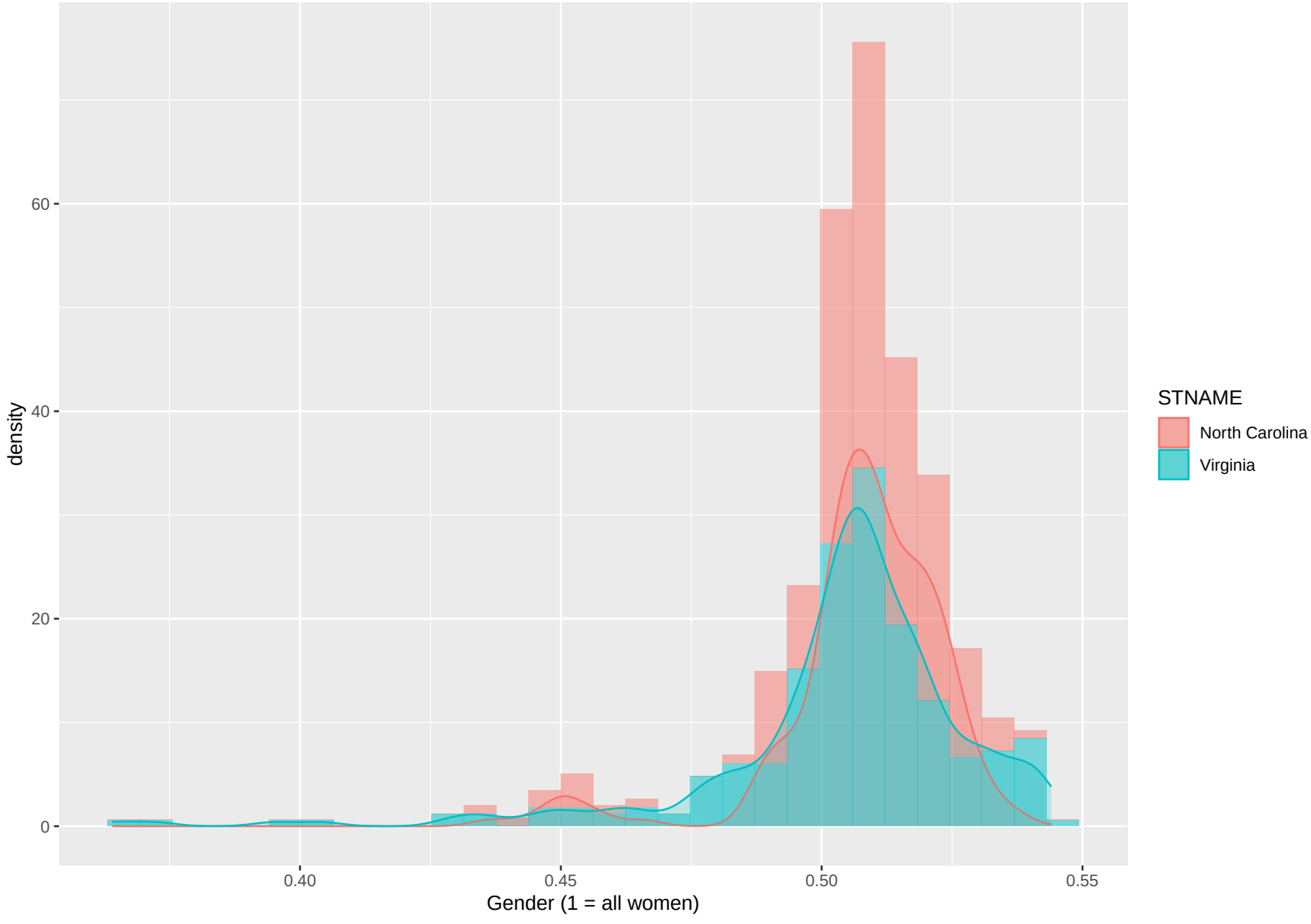
Density Plots

Per the results we would expect to see that Average Age, Gender, Percent White, and Percent Black are all very similar between the two states. Whereas Percent Hisp/Latino, Population Density, College Graduation Rate, and the Employment Level are different.

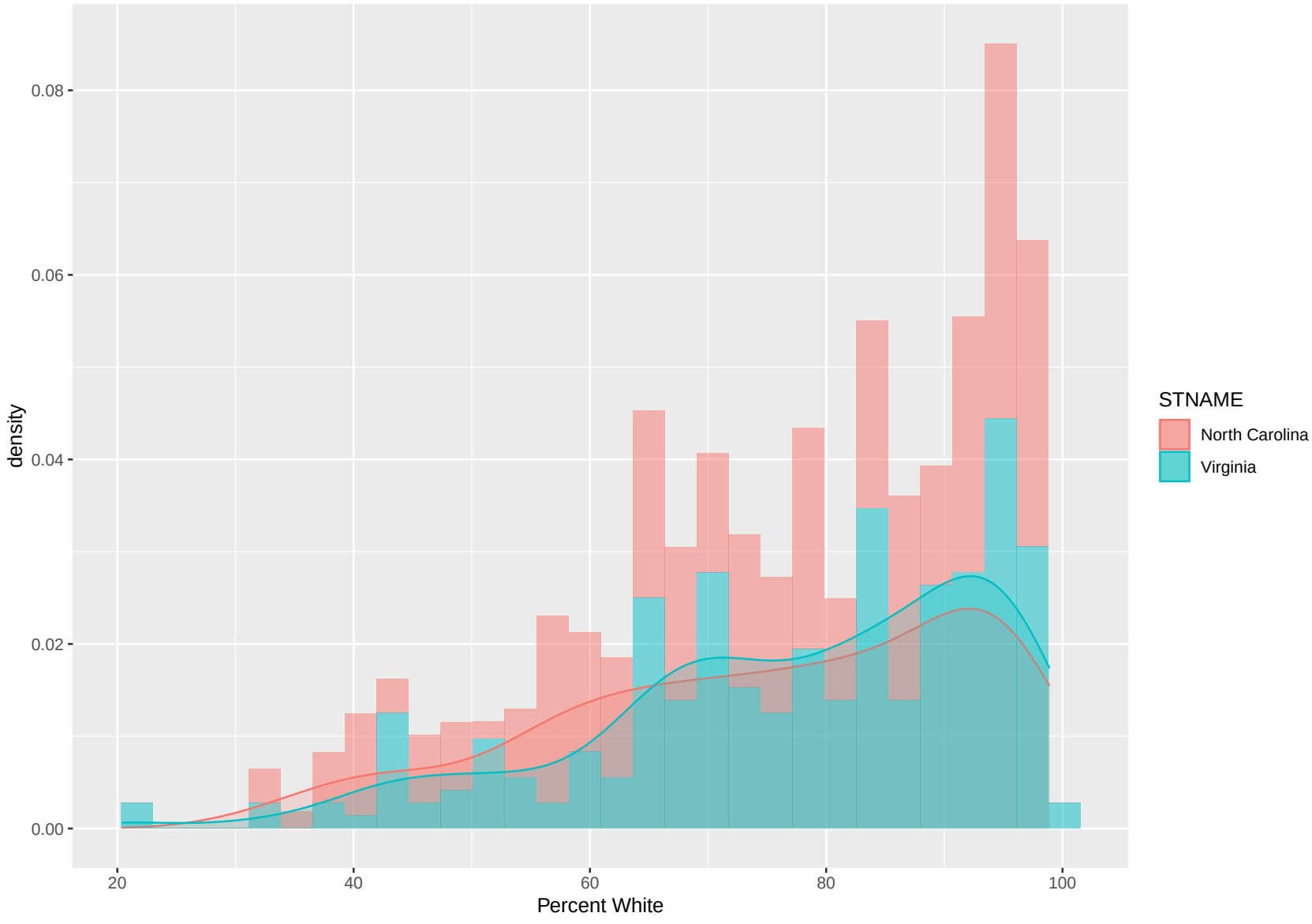
Age


1.5
1.0
0.5
0.0
density
7
8
9
10
11
Average Age
(9 = 40−44, 10 = 45−49)
STNAME
North Carolina
Virginia

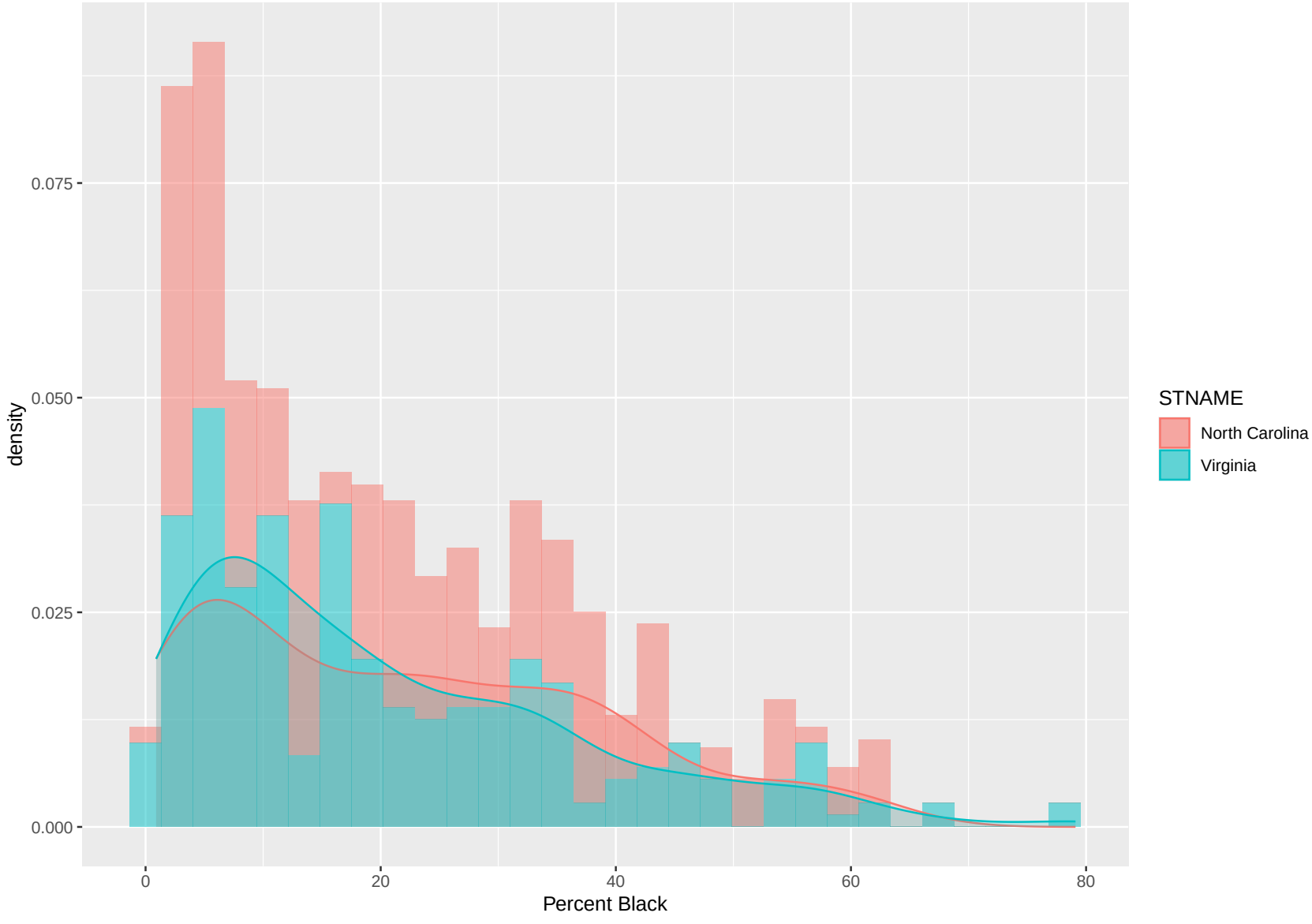
Gender


60
40
20
0
density
0.40
0.45
0.50
0.55
Gender (1 = all women)
STNAME
North Carolina
Virginia

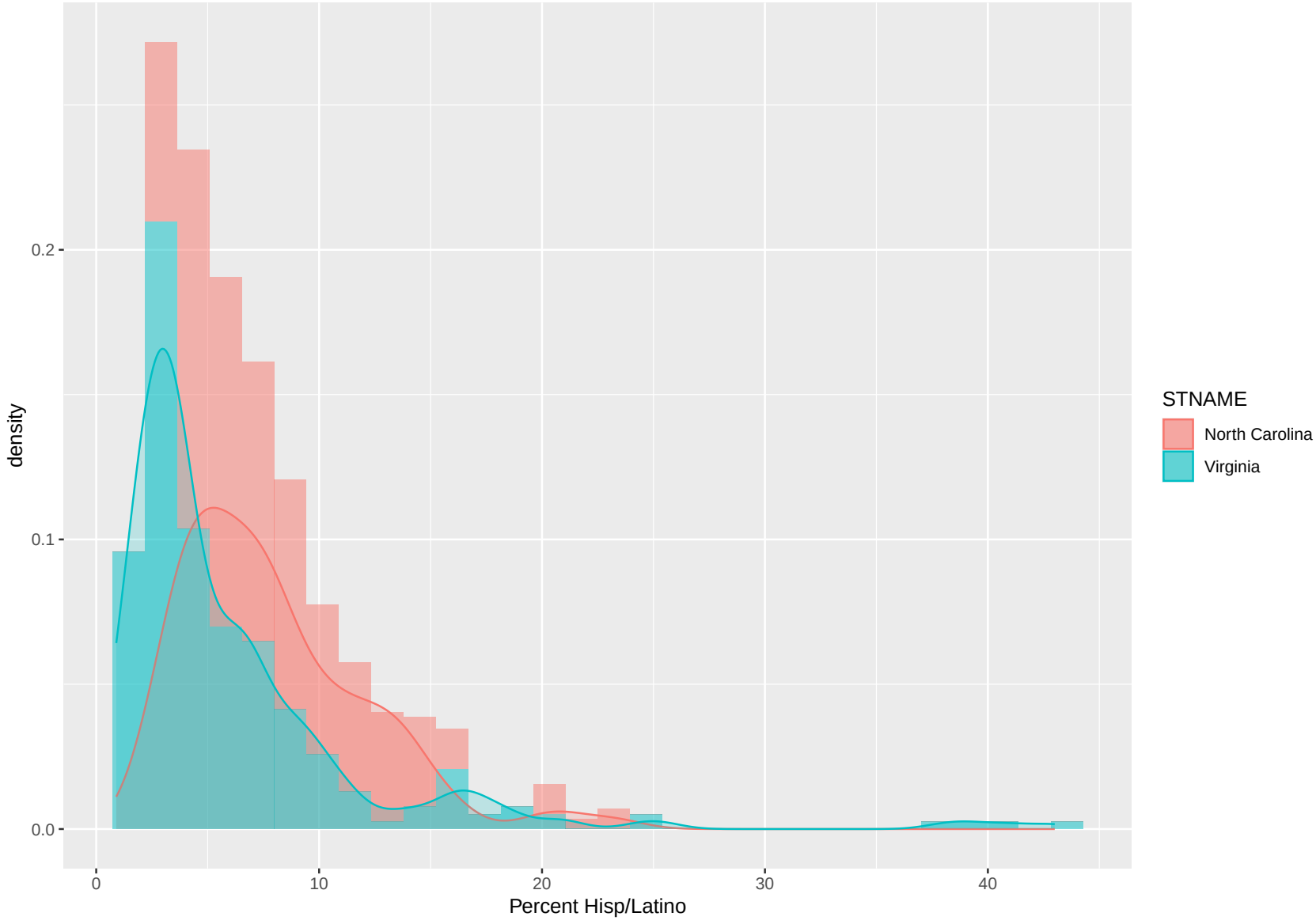
Percent White



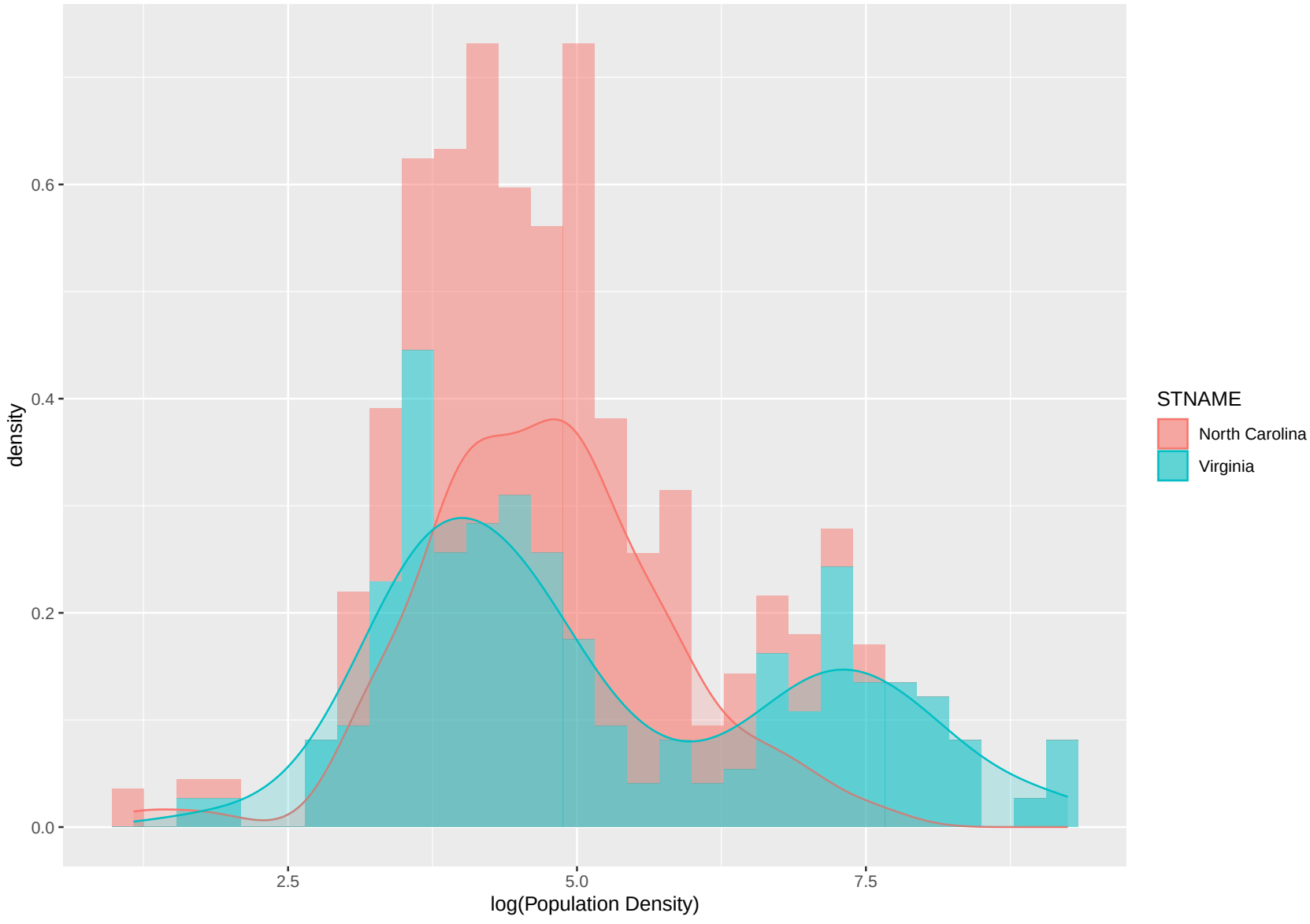
Percent Black



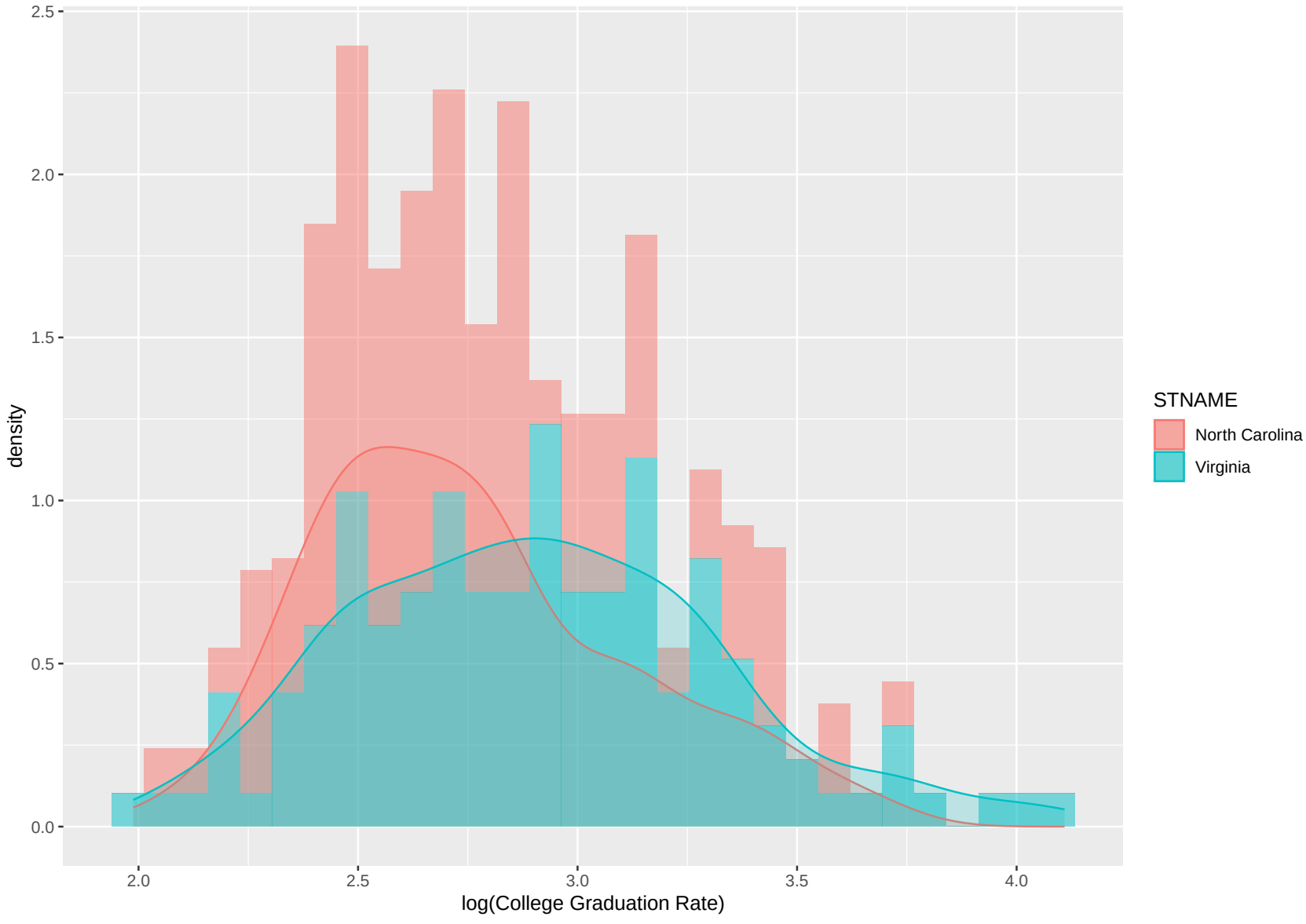
Percent Hispanic



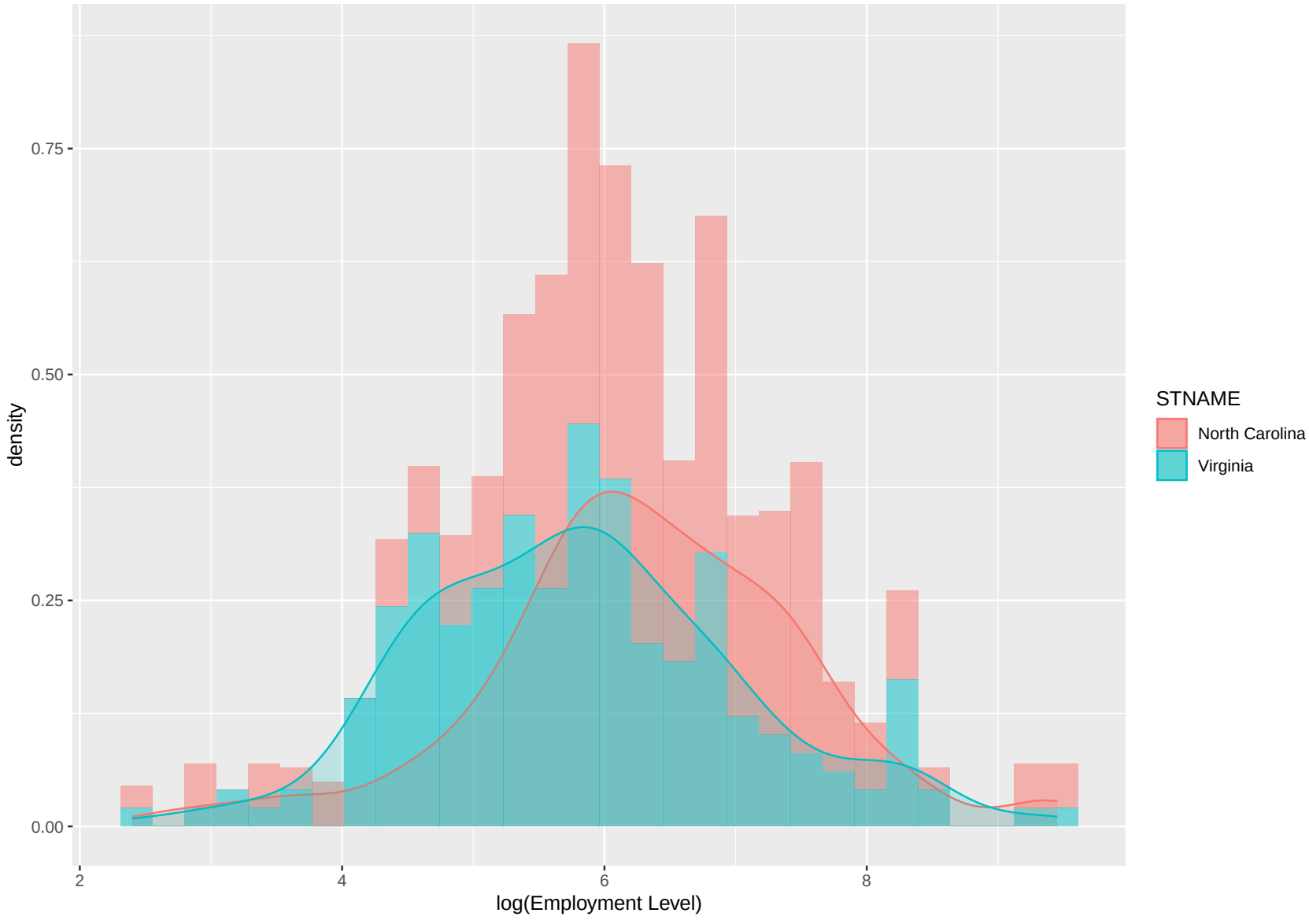
Population Density



College Graduation Rate



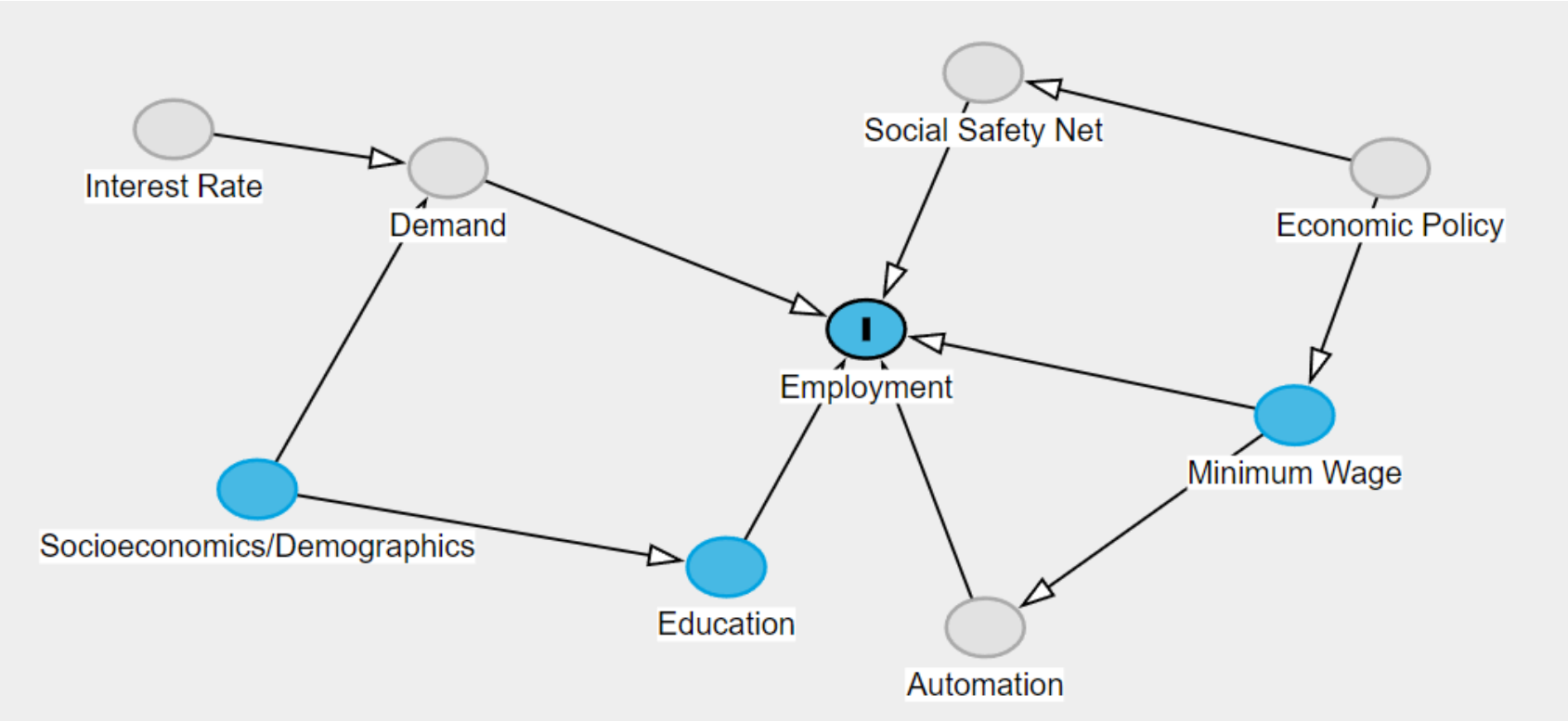
Employment Level



Research Design

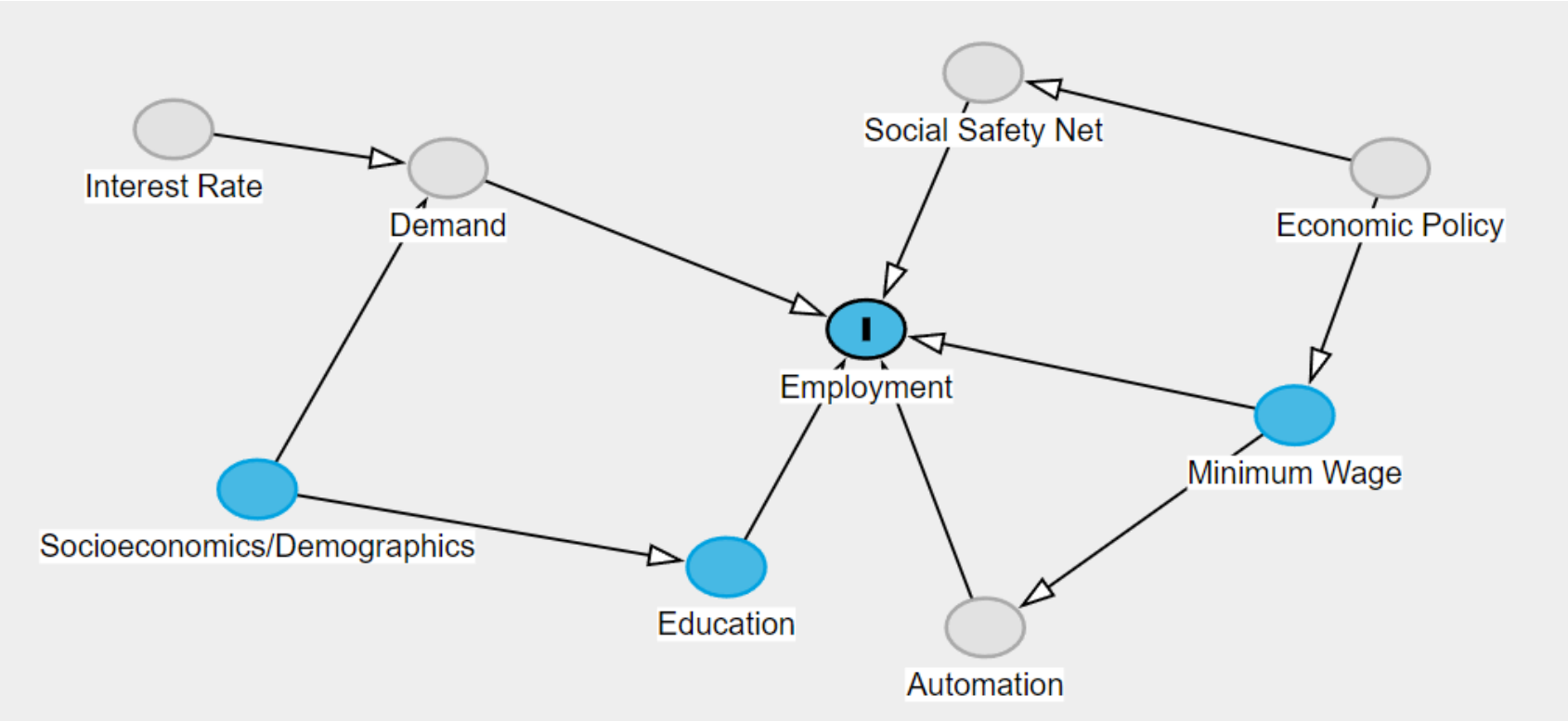
Directional Acyclic Graph

- The DAG contains my research assumptions for the major roots of the employment level in a typical western capitalist economy.
- Grey are roots whose data is not available to me or are too abstract to include in a model
- Blue are roots that I had data for and are accounted for in the model



Research Design
Model Design

- In order to run an appropriate regression all direct effects need to be controlled for if possible
- This means education needs to be controlled for
- It also means demographics needs to be included since it is a confounder through demand
- In this way demographics will act as a proxy for demand



Research Context

Criticisms of Difference in Differences

- A 2019 paper by Bruno Ferman and Cristine Pinto was published criticizing DID as a method
- They challenged the prevailing assumption that having a large number of observations for each group satisfies the homoskedasticity requirement for multilinear regression
- They also suggested that if the treated group is much larger than the control, then DID would tend to under-reject the null hypothesis

Research Context

Criticisms of Difference in Differences

- In both my paper and the Card-Krueger paper, the treated group is larger than the control, but mine is much closer:
  - 331 NJ and 79 PA
  - 94 VA and 79 NC

Research Context

DID Assumptions

- Intervention unrelated to outcome at baseline (allocation of intervention was not determined by outcome)
- Treatment/intervention and control groups have Parallel Trends in outcome
- Composition of intervention and comparison groups is stable for repeated cross-sectional design
- No spillover effects

Research Context

Lasso Regression

$$\arg\min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2\right\}+\lambda\sum_{j=1}^{p}|\beta_j|$$

$$E(\beta)+\lambda R(\beta)$$

Data

Selection Bias Problem

Because of the large number of missing counties in the employment level data, selection bias may influence the results. Counties that do not have the resources to collect employment level data may be more rural with a lower average income and may be more exposed to a minimum wage hike.

Data

Bureau of Labor Statistics

- Employment level data for food and beverage retailers (industry code: 445)
  - 2019 Q1
  - 2021 Q4
  - 2022 Q1

US Census

- County Land Area
- Number of people with university degrees
  - US Census American Community Survey on Educational Attainment
  - 2020
- All Other Demographics
  - Calculated from County Population by Characteristics Dataset
  - 2019
  - 2021

Data

Table 1: Descriptive Statistics: Food Service Employment Level ($TotalVA = 133$ $TotalNC = 100$)

Subset	N	Mean	StdDev	Min	Max
VA 2019 Q1	105	716	1382	19	11091
VA 2021 Q4	98	805	1487	11	11776
VA 2022 Q1	107	745	1447	10	11988
NC 2019 Q1	84	1058	1818	20	11443
NC 2021 Q4	82	1175	2008	12	12746
NC 2022 Q1	84	1145	1993	10	12336

Data

Table 2: Descriptive Statistics: Education Data 2020 (*All233Counties*)

Measure	Mean	StdDev	Min	Max
Graduation Rate	18.64	8.49	7.3	60.9

Table 3: Descriptive Statistics: Data for Year 2019 (*All233Counties*)

Measure	Mean	StdDev	Min	Max
Average age (grouped)	8.92	0.69	6.93	10.86
Gender (1 = 100% women)	0.51	0.02	0.37	0.54
Percent White	76.80	17.09	20.60	98.90
Percent Black	21.00	16.62	0.90	78.90
Percent Hisp/Latino	6.67	5.35	0.90	41.00
Population Density	579.77	1380.34	3.50	10365.90

Data

Table 4: Descriptive Statistics: Data for Year 2021 (*All233Counties*)

Measure	Mean	StdDev	Min	Max
Average age (grouped)	8.93	0.70	6.96	10.80
Gender (1 = 100% women)	0.50	0.02	0.36	0.54
Percent White	76.64	17.03	20.40	98.80
Percent Black	21.08	16.52	1.00	79.10
Percent Hisp/Latino	7.10	5.50	1.00	43.00
Population Density	579.68	1358.68	3.20	10058.90

Data

The DID Dummies

- Control
  - 1 for VA
  - 0 for NC
- Time
  - 0 for Initial
  - 1 for Final
- Treatment
  - 1 for the VA after the treatment
  - 0 for all else

Table 6: Comparing the Original Regression to the Lassoed Regression

	Dependent variable:		
	Employment among Food and Beverage Retailers		
	Original	Lassoed	Control Test
	(1)	(2)	(3)
Time	103.932		62.737
	(212.090)		(148.002)
Control	−465.194**	−654.318****	
	(211.483)	(151.078)	
Minimum Wage Hike	−83.746		
	(285.939)		
Average age (grouped)	−477.285****	−431.529***	−367.700***
	(131.521)	(131.965)	(134.668)
Gender	−1,833.666		6.962
	(3,583.782)		(3,682.644)
Percent White	−13.036***	−10.526**	−11.244**
	(4.683)	(4.663)	(4.789)
Percent Hisp/Latino	30.304*	9.746	25.324
	(17.222)	(16.036)	(16.055)
Population Density	−0.240***		
	(0.075)		
College Graduation Rate	105.830****	87.167****	79.194****
	(10.351)	(8.630)	(8.849)
Constant	5,190.012**	4,150.095****	3,285.383
	(2,156.160)	(1,189.485)	(2,190.122)
Observations	369	369	369
R^2	0.349	0.329	0.295
Adjusted R^2	0.332	0.320	0.283
Residual Std. Error	1,364.857 (df = 359)	1,377.728 (df = 363)	1,414.474 (df = 362)
F Statistic	21.358**** (df = 9; 359)	35.595**** (df = 5; 363)	25.205**** (df = 6; 362)
Note:			* p<0.1; ** p<0.05; *** p<0.01; ****p<0.001

Variable Selection

Percent Hisp/Latino

- According to the Lasso threshold Percent Hisp/Latino should be a good predictor of employment, but after dropping the other weak variables from the model in Table 6, it actually gets significance shifted away from it
- Of the variables listed as highly significant ($p < .001$) in table 5, Percent Hisp/Latino is the only one still in our model
- This implies that it is only good at predicting employment insofar as it is good at differentiating between North Carolina and Virginia
- We should expect to see a large difference in concentrations of Hispanic populations in the two states

Variable Selection

Percent Hisp/Latino and Population Density are Good Classifiers

- In Table 7 I've run a logistic regression on Percent Hisp/Latino and Population Density compared to all the other demographics
- Hisp and Pop got an error rate of 27% and all the others got an error rate of 38%

Table 7: Hisp/Latino and Population Density are Good Classifiers

	Dependent variable:	
	Control: North Carolina = 0, Virginia = 1	
	Hisp and Population	Other
	(1)	(2)
Percent Hisp/Latino	−0.298****	
	(0.042)	
Population Density	0.003****	
	(0.0004)	
Average age (grouped)		0.005
		(0.172)
Gender		−13.921**
		(5.833)
Percent White		0.001
		(0.007)
College Graduation Rate		0.057****
		(0.015)
Constant	1.266****	6.031*
	(0.234)	(3.267)
Observations	369	369
Log Likelihood	−192.885	−243.257
Akaike Inf. Crit.	391.770	496.513
Note:	* p<0.1; ** p<0.05; *** p<0.01; ****p<0.001	

Variable Selection

Variables in Question: Gender

- Gender was not found to be significant in the initial regression
- Gender was shown to be a good classifier
- Gender was below the lambda threshold in the Lasso regression
- All three results imply Gender should be dropped in the final model

Variable Selection

Variables in Question: Percent Hisp/Latino

- Percent Hisp/Latino was not found to be significant in the initial regression ($p<.1$)
- Percent Hisp/Latino was shown to be good a classifier
- Percent Hisp/Latino is directly on the lambda threshold
- The first two items alone advise dropping it from the model and since its not definitively selected by the lasso as a good predictor it can be safely dropped

Variable Selection

Variables in Question: Population Density

- Population Density was found to be significant in the initial regression
- Population Density was shown to be a good classifier
- Population Density was below the lambda threshold in the Lasso regression
- These results imply that Population Density is only initially significant because it is proxying the difference of employment level between the two states