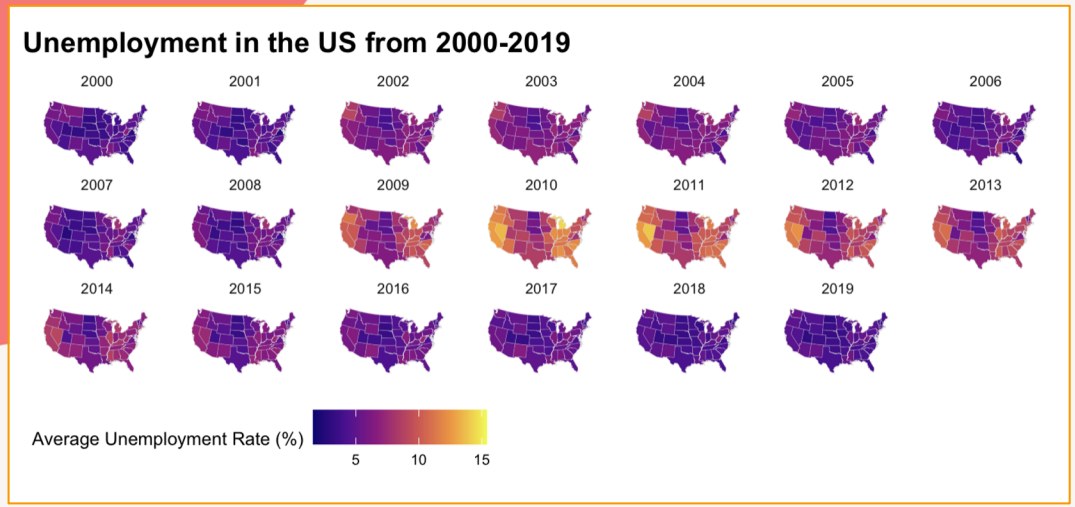
Unemployment In the United States

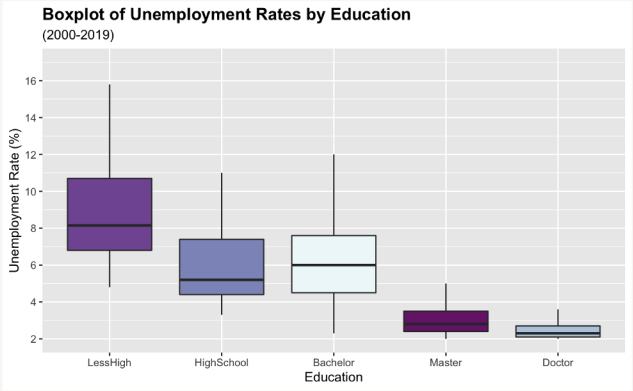

Political Party
Federal Surplus/Deficit
Household Income
Population
Gender
Education
Home Value Index
Personal Consumption Expenditure
Interest Rate
Race
We aim to explore and evaluate various economic, social, and geographic factors and their relationship with unemployment.

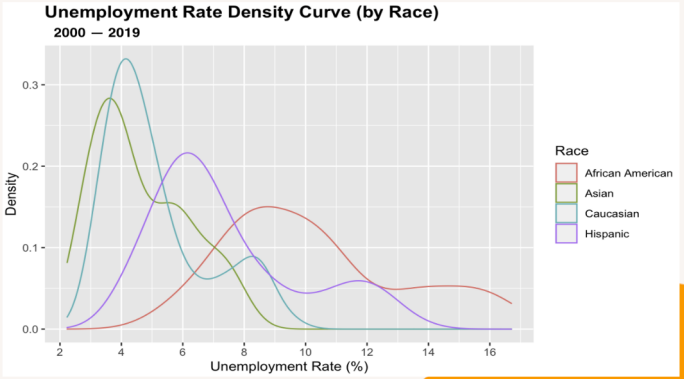
Findings



Multicollinearity

- ➔ Median Income vs. Home Value Index
- ➔ Home Value Index vs. Population
- ➔ Personal Consumption Expenditure vs. Federal Surplus/Deficit
- ➔ Political Party vs. Regions





Linear Regression model

Except Hispanic, other regressors are statistically significant in each simple regression model .
The p-values of statistically significant regressors close to 0. It means that the probability of getting the value as extreme as the respective t value is very small.
we could reject the null hypothesis because their respective p-values are <0.05. It indicates that with 95% confidence, we reject the null hypothesis and there is evidence to support the alternative hypothesis.

Figure : Simple Linear Regression

	β0	β1	P value	R^2
Woman	6.0896	-0.4408	.00981 **	0.01387
Asian	7.5921	-2.9325	2.4×10^{-14} ***	0.1618
Caucasian	7.4761	-2.3439	1.2×10^{-9} ***	0.1061
African American	5.7362	4.5329	2.2×10^{-16} ***	0.3969
Hispanic	6.7164	0.6586	0.0959 .	0.008378

Model1: Multiple linear regression

the relationship between a state's *Median Income* and *Unemployment Rate*

$$\textit{Unemployment Rate} = 0 + (8.22 \times 10^{-5})\ \textit{Median\_Income} + 0.57\ \textit{West} + 0.62\ \textit{Northeast} - 0.17\ \textit{Midwest} + 1.46\ \textit{Republican} + u$$

However, we noted that these results also show that there is no significant relationship between population and unemployment rate since the p-value is > 0.1, which leads us to model 2.

Model 1 – Multiple Linear Regression

	Dependent variable:
	Rate
Median_Income	0.0001***
	(0.00000)
West	0.572***
	(0.059)
Northeast	0.616***
	(0.068)
Midwest	−0.166***
	(0.060)
Republican	1.464***
	(0.046)
Observations	12,000
R2	0.826
Adjusted R2	0.826
Residual Std. Error	2.443
F Statistic	11,410.000***
Note:	*p<0.1; **p<0.05; ***p<0.01

Model2: Multiple linear regression

$$\textit{Unemployment Rate} = 0 + 0.35 \log(\textit{Median\_Income}) + 0.73\ \textit{West} + 0.36\ \textit{Northeast} + 0.36\ \textit{Midwest} - 0.16\ \textit{Republican} + 0.05\ \textit{AboveHigh\_Rank} + (4.56 \times 10^{-9})\ \textit{Population} + u$$

Based on our regression results in Figure, we found that the adjusted R2 value has increased to 0.889 and the F-statistics value increased to 13,770. The residual standard error of the model has decreased and remains relatively small with a value of 1.95. This implies that the new model is a better fit. However, we noted that these results also show that there is no significant relationship between population and unemployment rate since the p-value is > 0.1.

Model 2 – Multiple Linear Regression

	Dependent variable:
	Rate
log(Median_Income)	0.353***
	(0.007)
West	0.732***
	(0.059)
Northeast	0.356***
	(0.068)
Midwest	0.362***
	(0.061)
Republican	−0.157***
	(0.044)
Population	0.000
	(0.000)
AboveHigh_Rank	0.054***
	(0.002)
Observations	12,000
R2	0.889
Adjusted R2	0.889
Residual Std. Error	1.950
F Statistic	13,769.820***
Note:	*p<0.1; **p<0.05; ***p<0.01

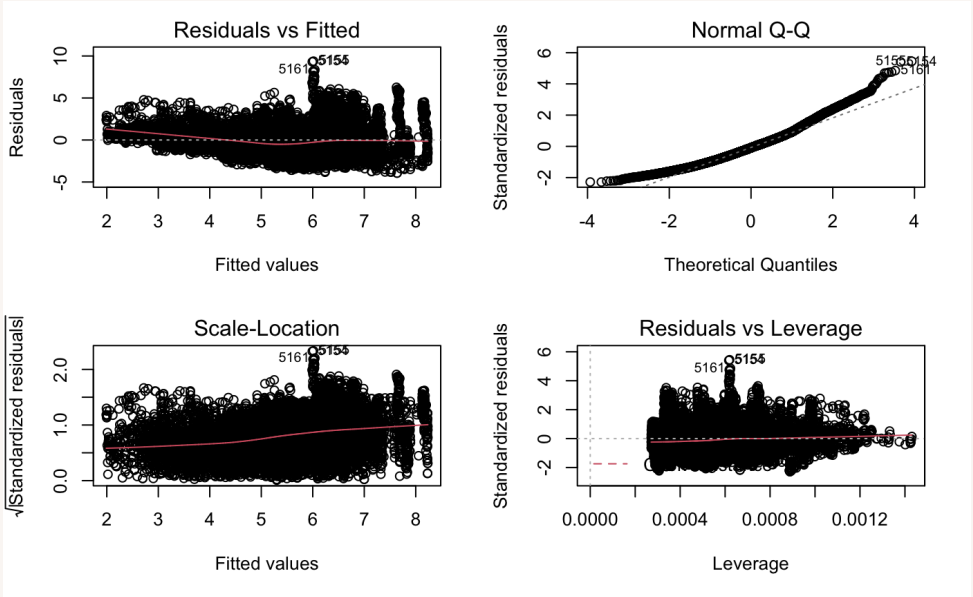
Multicollinearity Diagnostics

A multicollinearity diagnostics analysis of the variance inflation factor, the tolerance and Farrar-Glauber F-test, we identified that there is multicollinearity located in the model as shown in Figure.

```
Call:
imcdiag(mod = model1)


All Individual Multicollinearity Diagnostics Result

                VIF    TOL        Wi       Fi Leamer   CVIF Klein  IND1   IND2
West         1.9504 0.5127 2279.7029 2849.866 0.7161 2.0234     1 2e-04 1.0835
Northeast    2.2950 0.4357 3106.4952 3883.443 0.6601 2.3810     1 2e-04 1.2548
Midwest      2.0489 0.4881 2516.0934 3145.379 0.6986 2.1256     1 2e-04 1.1384
Republican   1.5355 0.6513 1284.4993 1605.758 0.8070 1.5930     1 3e-04 0.7755
Population   1.3631 0.7336  871.0549 1088.909 0.8565 1.4142     1 3e-04 0.5924
AboveHigh_Rank 2.0816 0.4804 2594.5697 3243.483 0.6931 2.1596     1 2e-04 1.1554

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

* all coefficients have significant t-ratios

R-square of y on all x: 0.1083

* use method argument to check which regressors may be the reason of collinearity
===================================
```

Heteroskedasticity test

A 1% increase in a state's median income is associated with a 0.042% increase in unemployment rate, and a 1% increase in interest rate is associated with a -0.475% decrease in unemployment rate



Confidence Intervals

Multicollinearity between *West*, *Northeast*, *Midwest*, *Republican* and *AboveHigh_Rank* in figure above,but R^2 value of y on all x of 0.3183 to be reasonable and appropriate.

Thus, we reject the null hypothesis at the 95% confidence level.

	2.5 %	97.5 %
log(Median_Income)	0.40981545	0.43454906
Interest_Rate	-0.49140472	-0.45931940
West	0.68780991	0.88286922
Northeast	0.29890699	0.54126953
Midwest	0.32370384	0.52649690
Republican	-0.20886550	-0.05954941
AboveHigh_Rank	0.05400262	0.05930686

Logistic regression analysis

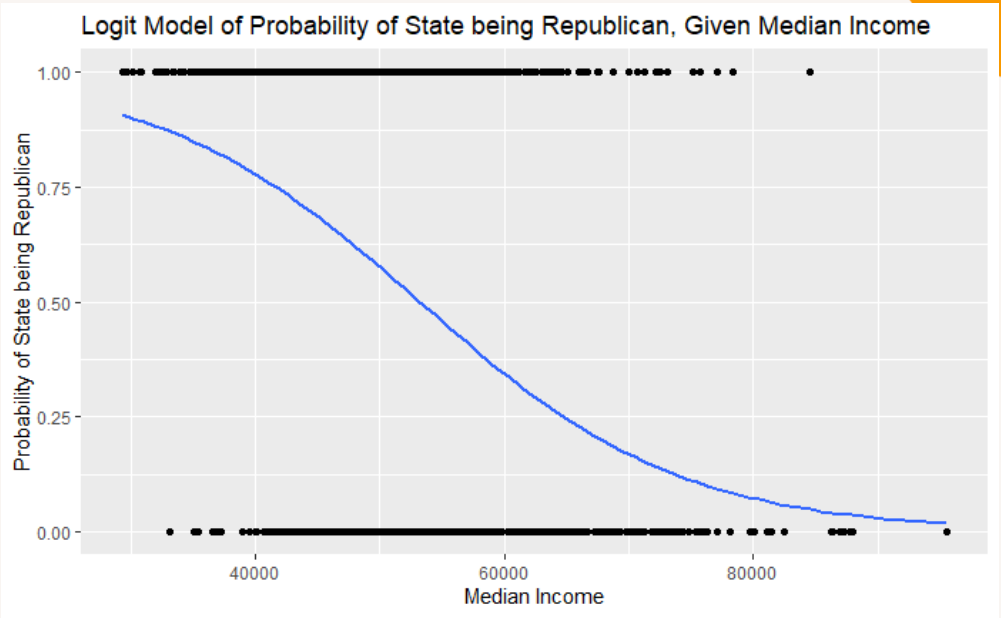
model the probability of a State having high unemployment.

- Pr (Y = High_Unemp = 1 | X) = F(β0 + β1 House Value Index)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.0353e-01	2.8601e-02	17.605	< 2.2e-16	***
HVI	-1.9266e-06	1.3484e-07	-14.288	< 2.2e-16	***

- Pr (Y = High_Unemp = 1 | X) = F(β0 + β1 Population + β2 West)

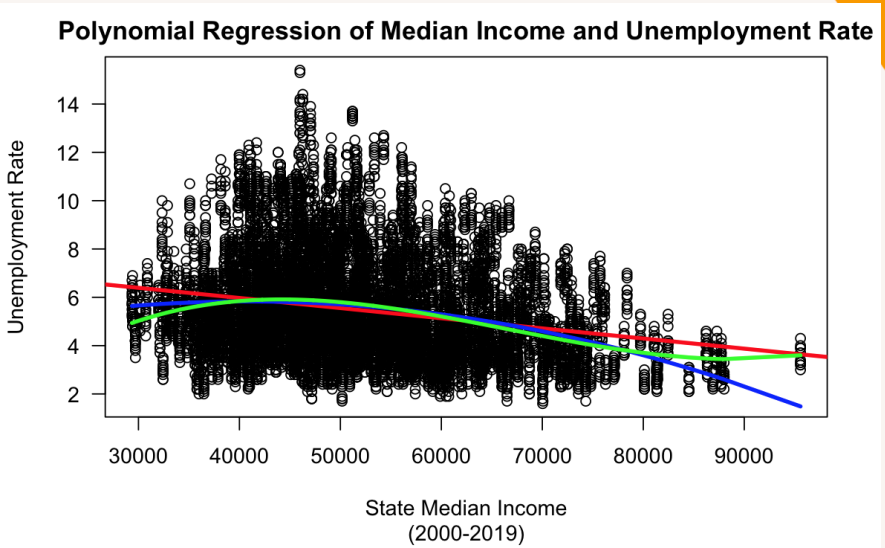
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.0556e-01	1.8583e-02	-5.6805	1.343e-08	***
Population	2.9398e-08	2.0663e-09	14.2273	< 2.2e-16	***
West	1.8224e-01	2.7019e-02	6.7449	1.532e-11	***

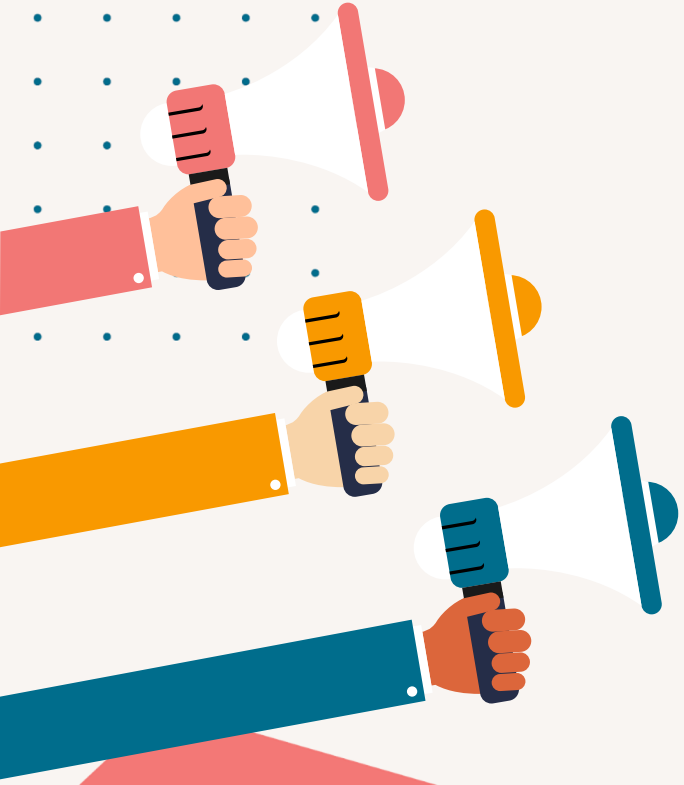


Polynomial Model

$$Unemployment\ Rate = -5.833 + 6.401 \times 10^{-4}(Income) - 1.084 \times 10^{-8}(Income^2) + 5.42 \times 10^{-14}(Income^3) + u$$

```
Call:
lm(formula = y ~ norm + square + cubic, data = state_model)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0916 -1.4060 -0.4219  1.0122  9.4993

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.883e+00  1.139e+00  -5.166 2.43e-07 ***
norm         6.401e-04  6.230e-05  10.275  < 2e-16 ***
square      -1.084e-08  1.104e-09  -9.817  < 2e-16 ***
cubic        5.420e-14  6.339e-15   8.550  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.98 on 11996 degrees of freedom
Multiple R-squared:  0.06896,   Adjusted R-squared:  0.06872
F-statistic: 296.2 on 3 and 11996 DF,  p-value: < 2.2e-16
```





THANK YOU!