Introduction

Team 10 has been hired by a national law firm looking to expand upon its practice by opening up a regional or state-specific office with the primary focus on workplace discrimination lawsuit resolution. The firm has enough capital and resources to open up a law firm of any size, anywhere within the country but lacks an understanding of where would be the best location. They want to open up a location that will have the opportunity to have an outsized impact on the community that it serves, not necessarily just a location that handles the most cases.

Problem Statement

<u>The problem statement for the firm is that they lack a fundamental understanding of the current landscape of discriminatory practices throughout the nation and do not know where would be the best location to open up their first discrimination practice office</u>.

The firm may have the right business acumen and talent when it comes to actually running a law firm, but they need team 10 to help guide them in terms of identifying where an outsized market opportunity exists for them.

Objectives

They have multiple competitors within the nation already with established discrimination practices and want to make sure that the first office that they open up is in the optimal location. From an economic standpoint, the office should be opened in a location that has a large volume of discriminatory lawsuits being filed in order to support the firm financially. This does not mean that the firm wants to open up a location that handles the most volume, but rather that there is enough volume to support their venture. From a talent standpoint, the law firm would also like to understand whether or not they should be hiring lawyers with a specific law specialty to support their office opening. From a corporate social responsibility standpoint, leadership at the firm is suspicious that there may be certain regions and states within the country that have suppressed volumes of discrimination being filed & would like to understand whether or not opening up an office in that region or state would make sense for a secondary opening. Additionally, as noted above, the firm would be interested in making an outsized social impact with its first opening. Put a different way, leadership has shared with us that one of its objectives is to make sure that the community that it initially serves has an outsized volume of discriminatory lawsuits.

To expand upon the talent question - there are multiple different discrimination laws within the United States. The firm is keenly aware of this and wants to make sure that its talent hiring for the new location is appropriately tailored to the volumes they anticipate working on. They also would love to understand what states have abnormally high proportions of certain discriminatory laws being filed - to understand whether or not they should instead focus on specific subsections of discriminatory law in the United States.

<u>To summarize the team 10's objectives as they relate to the law firm:</u>

1. To identify the most attractive region or state to open up an office, both in the current and near term. This could be based on the level of disproportionate claims being filed in one state.
2. To identify what specific discrimination law will be the primary focus of this office.

3. To identify any states or regions in which discriminatory lawsuits are abnormally low as targets for secondary office openings.
4. (If possible) Quantify the impact of the implemented state legislature to identify proactive opportunities for secondary office openings.

As trusted consultants to the national law firm in question, team 10 has been given quite a large amount of autonomy in determining the best way to answer the questions. We will approach the question from multiple different avenues, to account for the broad scope of the question that we have been hired to solve.

Description of Datasets

We aggregated data from many different sources in order to get a representative demographic of each state.

FY 2009 - 2021 EEOC Charge Receipts

Our main dataset came from the U.S. Equal Employment Opportunity Commission. Here we have yearly data for each state, grouped by type of complaint:

- Race
- Sex
- National Origin
- Religion
- Color
- Retaliation
- Age
- Disability
- Equal Pay Act
- Genetic Information Discrimination in Employment

From this data we were able to extract which states had the highest/lowest reports of workplace discrimination. However, we were unable to make useful regressions for factors that might influence these differences. We decided to look at two other types of metadata about each state: ethnic demographic, incumbent state governor and state's per capita income.

Bridged-Race Postcensal Population Estimates

The CDC collected population estimates for different demographics from 2010-2020. They collect on 10 different groups:

- White male/female
- Black male/female
- Native American male/female
- Asian male/female
- Hispanic male/female

From this data, we are able to see if states/regions with different racial make-ups have different reports of discrimination.

National Governors Association

The National Governors Association (NGA) has data on the incumbent governor for each state every year between 2010-2020. We wanted to see if the political climate in a state might influence the rate of discrimination reported.

Bureau of Economic Analysis

We incorporated per capita income of every state from the Bureau of Economic Analysis (BEA) to explore if socio-economic factors have an impact on the number of discrimination complaints filed in each state.

Population Data

Although the EEOC discrimination dataset has number of cases reported, it does not have the total population for the state. We appended this census data to our dataset in order to normalize the number of cases.

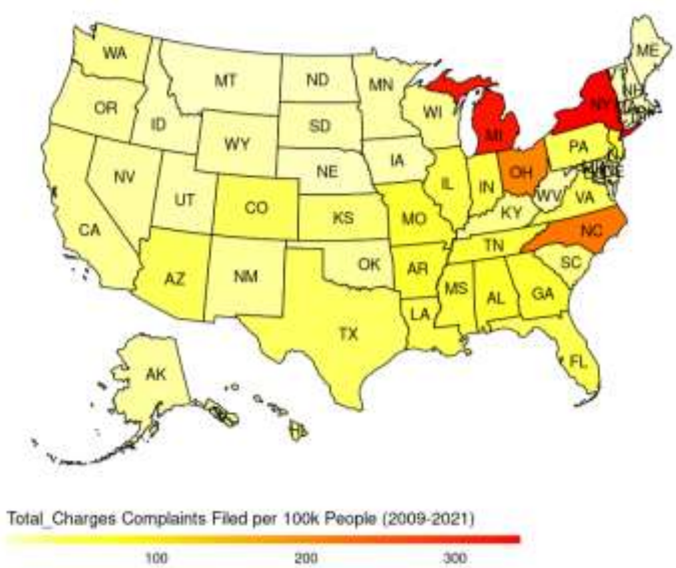
Approach and Methodology

Exploratory Data Analysis

State Analysis

Prior to data modeling, we wanted to get a sense of the types of correlation that may exist in our data. To start off, we aggregated total charges by state and examined which states had the largest number of complaints filed in 13 years.

Figure 1. Total Number of Complaints Filed per 100k Residents (2009-2021)



It is interesting to see that NY and MI appear to have the highest number of total charges filed per 100,000 people. We wanted to explore if any changes exist depending on the type of charge, so we further filtered the data by complaint type.

The full output with all 12 categories is in the appendix, but two interesting filters are "Race" and "National Origin" complaints. We see that Michigan and North Carolina have fairly low counts of "National Origin" complaints, however, they have large reports of race-based complaints. It is also interesting to note that New York is consistently the top reporter. This made us decide to look deeper into New York City using time series data.

Figure 2. National Origin and Race Complaints Filed per 100k Residents (2009-2021)

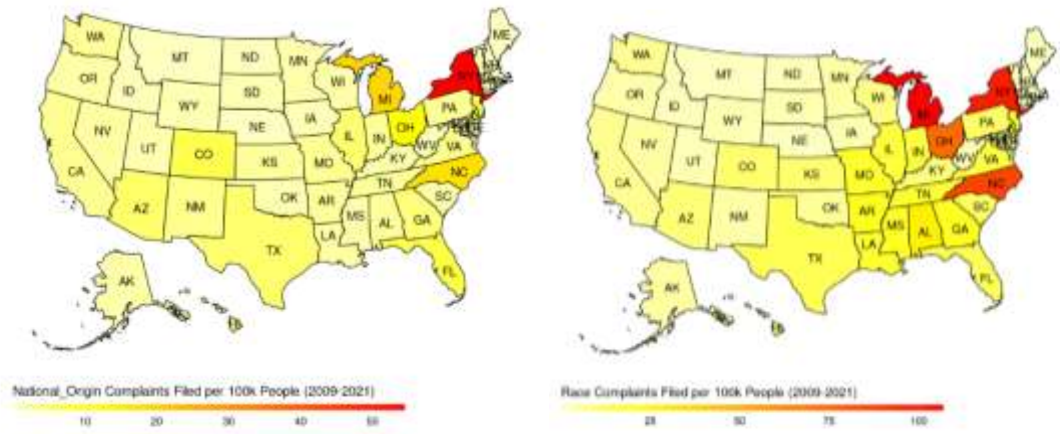


Figure 3. Top 3 States by Charge Category (per 100k Residents)

State	Age
New York	1,026.72
Michigan	950.72
Ohio	726.98

State	Race
Michigan	1,399.03
New York	1,287.76
North Carolina	1,113.47

State	Equal Pay
New York	58.65
Michigan	41.90
Ohio	33.73

State	National origin
New York	713.11
Michigan	249.83
North Carolina	223.78

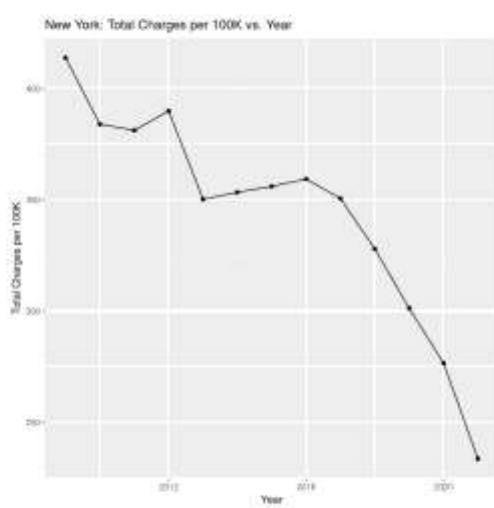
State	Religion
New York	242.07
Michigan	142.58
North Carolina	112.22

State	Sex
New York	1,455.53
Michigan	1,170.37
North Carolina	815.21

Time Series Analysis

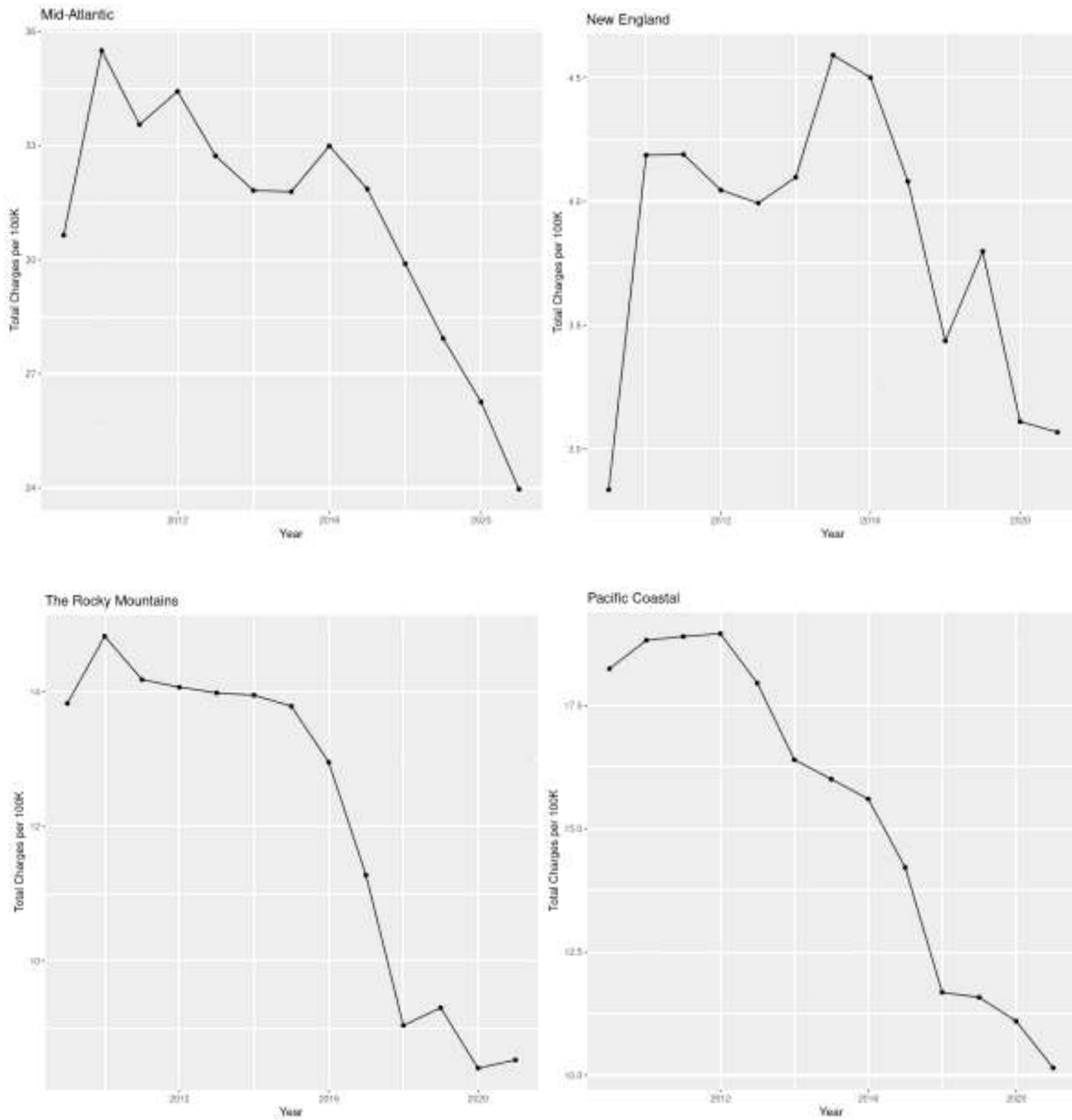
We wanted to look at the state of New York's total reported cases from 2009-2021. Here we see a sharp decrease in the number of cases reported after 2016.

Figure 4. Time Series Graph of Total Charges Filed in New York State from 2009-2021



In order to know if this was a localized anomaly or a global trend, we plotted all of the states in the Mid-Atlantic region, along with some other regions (to view all of the geographical regions' time-series graphs please view the appendix).

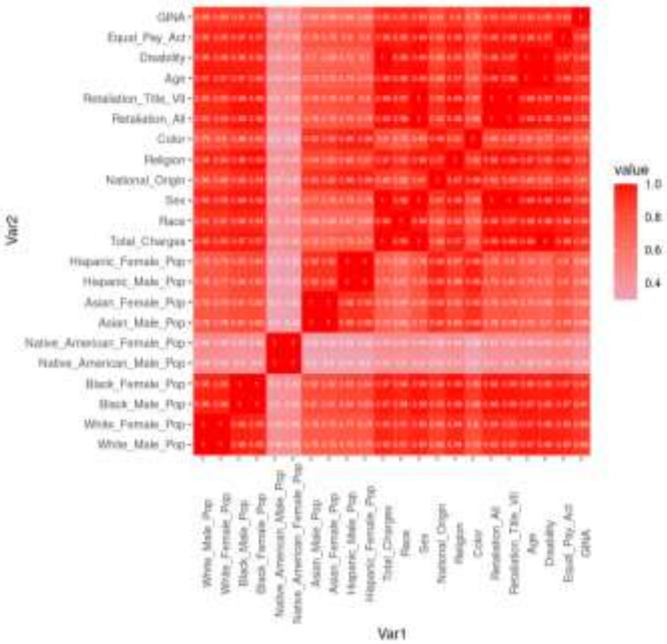
Figure 5. Time Series Graphs of Total Complaints Filed by US Regions from 2009-2021



While New York closely follows the Mid-Atlantic trend line, it is interesting to note that the New England trend line is not consistent. While the Rockies, Pacific, and Mid Atlantic all show a constant decrease, New England has a spike around 2018. We keep this in mind as we continue looking for factors that could influence certain types of discrimination reporting.

Correlation Heat Map

Figure 6. Heat Correlation Map of Demographics and Charge Category



We brought in more demographic data (gender and race) to see if there is a correlation with the types of reports and these factors. To visualize, we made a correlation heatmap (see appendix for full size). Unsurprisingly, many of these factors are highly correlated. Race is completely correlated between genders (if there is a large "White_Male_Pop" there is going to be a large "White_Female_Pop"). It is also interesting that a lot of complaint types are correlated as well, although this is confirmed by our U.S. map of complaint types by state. We've also observed that Native Americans demographics did not demonstrate correlation. Some further research showed that Native Americans have very low participation rates in the workforce, so it could be partially attributed to that (participation rate in workforce for example was 0.6% in 2013). However, legal standing of EEOC with the tribes is complicated. There have been cases where court rules that certain tribes are immune from discrimination law, which probably not just prohibits that tribe reporting but also could be discouraging other tribes' from filing complaints.

Regression Analysis

Linear Regression

To understand how the demographics of a region impact the total number of workplace discrimination complaints filed, we combined Bridged-Race Postcensal Population Estimates from the National Center for Health Statistics (NCHS) to our existing data. We also incorporated party affiliation information of the state governor for each year from the National Governors Association. We then built a linear regression model using this dataset to get a high-level view of any possible dependencies of the number of complaints filed with the EEOC on state demographics, leading party affiliations and state population makeup.

Model 1: The first model we developed, regressed Total_Charges (total number of discrimination charges filed with the EEOC for a given year) on State, Year, State_Population, Gov_Party (political party affiliation of the state governor for each year) and population estimates of White, Black, Native American, Asian and Hispanic male and female residents for a given year.

```
model1 <- lm(Total_Charges ~ Year + State + Gov_Party + State_Population +
White_Male_Pop +  White_Female_Pop + Black_Male_Pop + Black_Female_Pop +
Native_American_Male_Pop + Native_American_Female_Pop + Asian_Male_Pop +
Asian_Female_Pop + Hispanic_Male_Pop + Hispanic_Female_Pop, data)
```

The summary of the model (Output 1) showed that the population estimates of White, Native American, Asian and Hispanic residents of a state are statistically significant in explaining the number of complaints filed with the EEOC. However, the state governor's party affiliation did not seem to add any value in terms of explaining the response.

Model 2: We kept the demographic factors in order to explain our response and explored other variables which may have an impact on the number of complaints filed. From the US Bureau of Economic Analysis, we pulled each state's per capita income for a given year and combined it with our dataset. This enabled us to introduce an additional explanatory State_Per_Capita_Income to our linear regression model.

```
model2<-lm(Total_Charges ~ Year + State + State_Population +
State_Per_Capita_Income + White_Male_Pop + White_Female_Pop +
Black_Male_Pop + Black_Female_Pop + Native_American_Male_Pop +
Native_American_Female_Pop + Asian_Male_Pop + Asian_Female_Pop +
Hispanic_Male_Pop + Hispanic_Female_Pop, data)
```

The new factor (State_Per_Capita_Income) did not seem to add any significant explanatory value to our model (Output 2). Thus, we excluded that and kept the demographics data to further explore relationships between the population composition of a state with the number of discrimination complaints filed in it.

Model 3: In the third model, we used the demographic variables alone to explain the response. This led to Total_Charges being regressed over Year, State, State_Population and the population estimates of White, Black, Native American, Asian and Hispanic male and female residents for a given year.

```
model3 <- lm(Total_Charges ~ Year + State +  State_Population +
White_Male_Pop + White_Female_Pop + Black_Male_Pop + Black_Female_Pop +
Native_American_Male_Pop + Native_American_Female_Pop + Asian_Male_Pop +
Asian_Female_Pop + Hispanic_Male_Pop + Hispanic_Female_Pop, data)
```

This model showed that the variables selected are statistically significant (Output 3). However, to evaluate the performance of this model on our dataset, we measured how well the predictions made by this model matched the observed data. For this purpose, we used 10-fold cross-validation and compared its Root Mean Squared Error, R-Squared, Adjusted R-Squared and Mean Absolute Error with those of the previous two models (Table 1).

Table 1. Linear Regression Models Comparison

	R-Squared	*Adj. R-Squared*	*RMSE*	*MAE*
model1_cv	0.9854	0.985018	239.278	164.3648
model2_cv	0.9856	0.985305	238.521	162.4615
model3_cv	0.9860	0.985696	237.051	161.9039

Thus, we have a quantitative understanding of the performance of each model and found model3 to have the highest R-Squared and Adjusted R-Squared and the lowest RMSE and MAE after cross-validation. This justifies the use of this model to explain the number of charges one can expect in a given state.

Model 4: In an additional deep dive, we conducted a regression model on the racial charges related to the POC (people of color) population distributions. We conducted a separate analysis on Racial Charges relating to White population distributions, but did not find that White_Male_Pop and White_Female_Pop were significant predictors.

```
model4 <- lm(Race ~ Black_Male_Pop + Black_Female_Pop +
Native_American_Male_Pop + Native_American_Female_Pop + Asian_Male_Pop +
Asian_Female_Pop + Hispanic_Male_Pop + Hispanic_Female_Pop, data)
```

According to this model with R-squared adjusted value of 0.8853182, Black, Asian, and Hispanic Populations (both Male and Female) are significant predictors of racial charges, however they have coefficients that are very close to zero, and do not appear to have significant differences between them. This may imply that there are greater racial discrimination cases against these populations, because as population of these racial groups increases, so did the general number of racial charges.

Difference in Difference

To further understand the impact of specific state legislature that is being passed on discrimination filings in the United States, the team used the Difference-In-Difference estimation method. Specifically, we took a look at Missouri SB 43, which was passed in 2017. This bill made several changes to the existing “Missouri Human Rights Act”, which provides state law on many matters that overlap with the subject matter of the EEOC. This bill had several provisions that would theoretically have the impact of making proving discrimination within the workplace more difficult for plaintiffs as well as removing whistleblower and retaliation protections. The team would have expected that this bill would have the result of reducing discrimination filings within Missouri.

For illustrative purposes, several of the bill's provisions:

1) Limit the ability to fill class action lawsuits under state law in Missouri
2) Limit method of proving discrimination to be "the motivating factor", which is a higher threshold than previous
3) Remove various protections for whistleblowers as well as removing anti-retaliation provisions
4) Various adjustments to possibility punitive and actual awards

From a data massaging perspective, we created 4 distinct groups of reporting:

1) Control Group, Before Provision: Discriminatory filings in the US per capita before 2017
2) Control Group, After Provisions: Discriminatory filings in the US per capita after 2017
3) Treatment Group, Before Provision: Discriminatory filings in Missouri per capita before 2017
4) Treatment Group, After Provisions: Discriminatory filings in in Missouri per capita after 2017

We then created a linear regression model with:

Dependent Variable:

- Discriminatory Filings Per Capita

Independent Variables:

- Year (Before or After 2017)
- State (Missouri or otherwise)
- DiD (interaction variable of Year and State)

This model would theoretically allow us to understand whether or not Missouri has seen differentiated behavior regarding per capita discriminatory filings after the law was passed and implemented in 2017.

Results and Interpretation

Linear Regression

Our preliminary research indicated that socio-economic, demographic and political factors impact the number of complaints filed in each state. In order to better understand this relationship, we built 3 linear regression models which confirmed the following findings:

1) The year, state and its population were seen to be statistically significant in explaining the total number of charges filed in a state.
2) The region of a state was shown to not be statistically significant in explaining the total number of chargers filed in a state.
3) We hypothesized that political factors like state governor's party affiliation may have an impact on the number of complaints filed. But the regression results showed no consistent statistical significance of this variable.

4) Our research indicated that economic conditions like the state per capita income may play an important role in explaining the number of complaints filed. But the regression results showed no statistical significance of this variable.
5) In order to test if the population makeup of a state affects the number of discrimination complaints filed, we used population estimates of White, Black, Native American, Asian and Hispanic male and female residents for a given year. The regression results showed some of these variables to be statistically significant in explaining the total discrimination charges filed in a given state.

Difference in Difference

The Difference-in-Difference estimator resulted in the DiD coefficient being negative, as expected. This coefficient supports the premise that this SB 43 resulted in a slightly decreased discriminatory filings in the state of Missouri. From a data perspective, the coefficient and other factors around the model were not necessarily statistically robust, but, at the very least, the process itself will be valuable for the company.

Conclusion and Discussion

Our final recommendation to the law firm is as follows (corresponding to our original objectives):

1) Open up their first office(s) in:
    a) New York
    b) Michigan
    c) North Carolina
    d) Ohio
2) If they do decide to focus on one kind of law, focus on:
    a) Retaliation Discrimination
    b) Age Discrimination
3) If they wish to open up an office in a state that we believed it is possible that there is underreporting in:
    a) Minnesota
    b) Mississippi
4) Regarding state legislature:
    a) Proactively. The team believes that this can both be an opportunity and a threat for the business.

Recommendation #1 States to Open Offices in: Supported both by exploratory data analysis as well as linear regression modeling, we believe that the states of New York, Michigan, North Carolina & Ohio have outsized opportunities for the practice of discrimination law. Per Capita reporting was larger than other states in this subset of states. Additionally, the state coefficients associated with these states supports the idea that these states have some kind of underlying factor that leads to outsized discrimination reporting, all else considered equal. It is also worth noting that our linear regression modeling also concluded that regional reporting in the United States is not likely to be a statistically significant factor and thus, we do not recommend opening a regional office.

Recommendation #2 Law to Practice: Primarily derived from our exploratory data analysis, we believe that Age-related discrimination and retaliation-related discrimination represent the best opportunities for specialization of our client. These two types of law have outsized volumes in the states that we have recommended the firm considers opening up their office in.

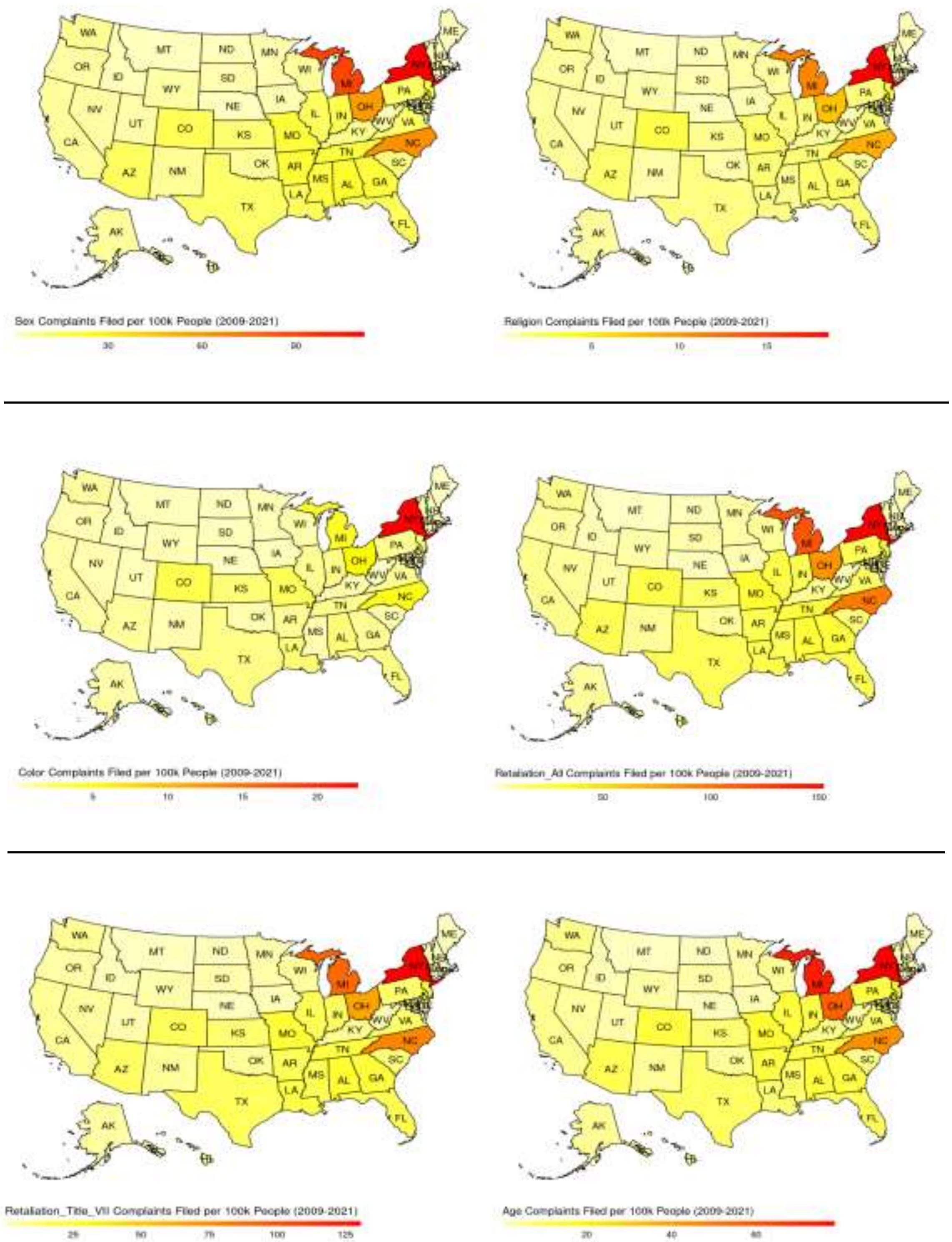
Recommendation #3 Additional States to Open Offices: Supported by our linear regression modeling, we believe that it is possible that both Minnesota and Mississippi have fewer than expected reporting of discrimination claims. This is based specifically upon the state coefficients developed in our models. Although it is difficult to surmise why this reporting is below the rest of the country, all else considered equal, it is worth investigating by the firm in the case that they believe that it is due to a lack of reporting. If this is the case, there could be untapped market share within the state.

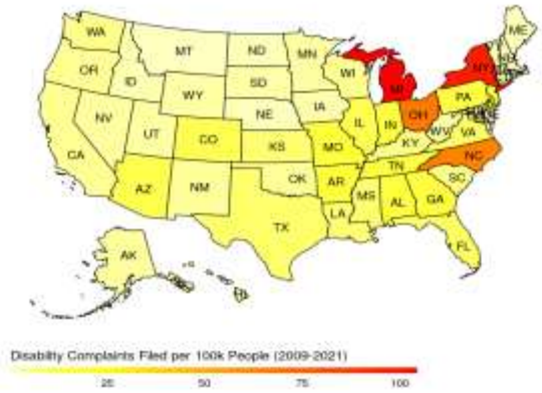
Recommendation #4 State Legislature: Supported primarily by our Difference-in-Difference estimator method - shifting states policies are likely to be an underlying cause of changes in the discriminatory filings within individual states, and thus, the value of a particular state as it relates to the law firms case volume. We therefore recommend that the law firm either identifies and engages in states with active laws that remove frictions from filing discriminatory claims or avoids states that are currently debating or have implemented laws that hinder the likelihood of an individual for filing discriminatory claims.

These laws' impacts are likely to be difficult to quantify, but the trend should be relatively easy to identify for a seasoned team of lawyers with specialization in the very subject matter the laws are impacting. Our recommendation is that the team works to understand the qualitative impacts of the political climate within individual states and work to be cognizant of that impact when making decisions to enter, leave or stay in a state. As indicated in our other work, differentiation in state-by-state discriminatory practices exists within the United States and the team believes that the state laws implemented are an underlying factor.

Appendix

Figure 7. Charge Category by State per 100k Residents (2009-2021)





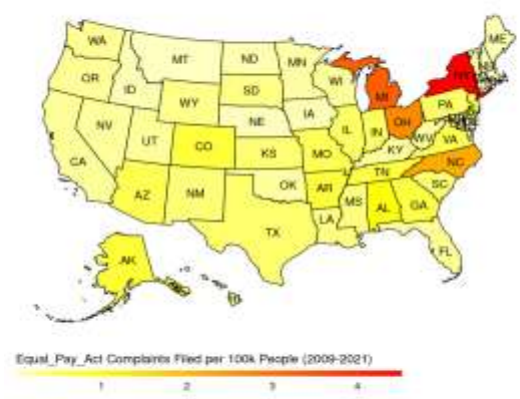
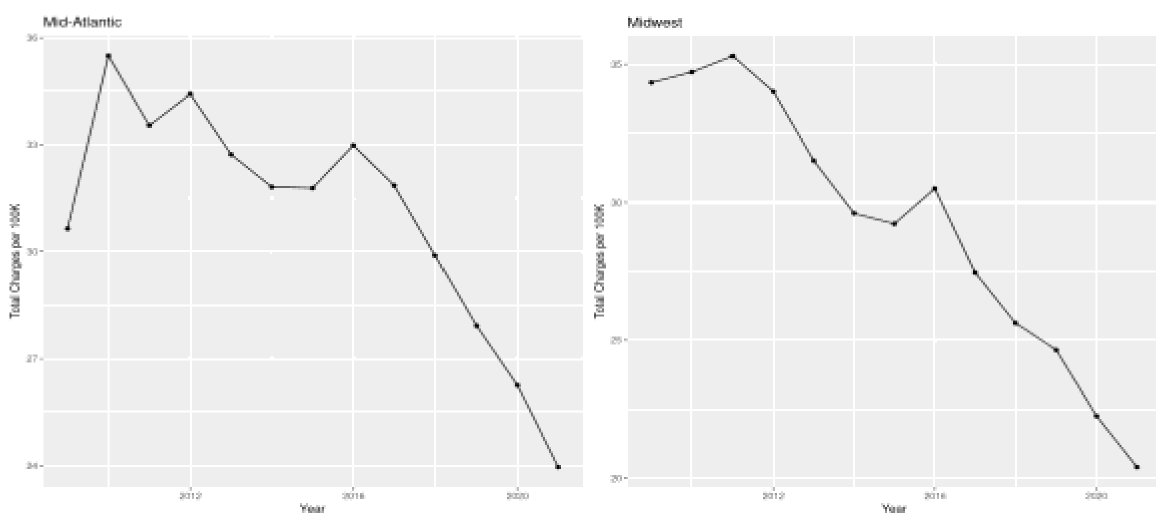
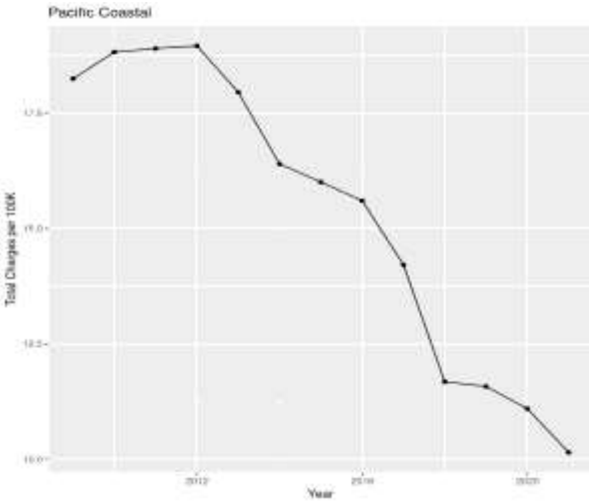
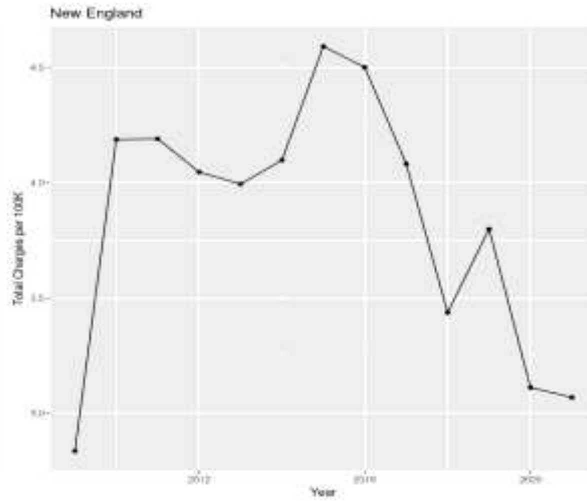
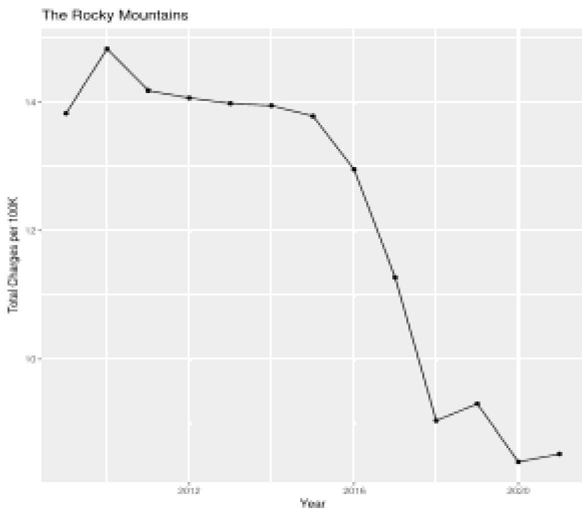


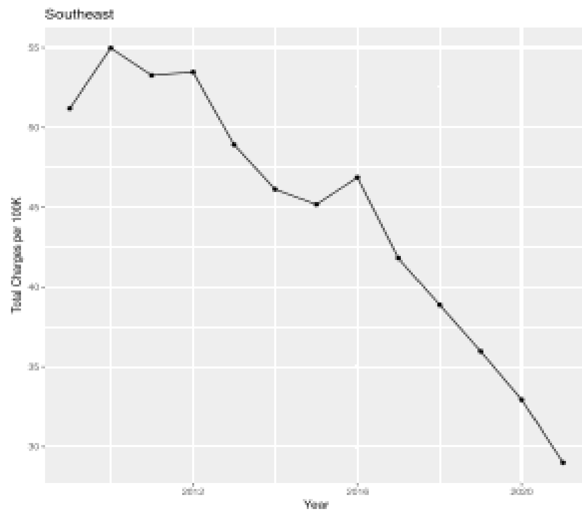
Figure 8. Time Series Graphs of Total Complaints Filed by US Regions from 2009-2021

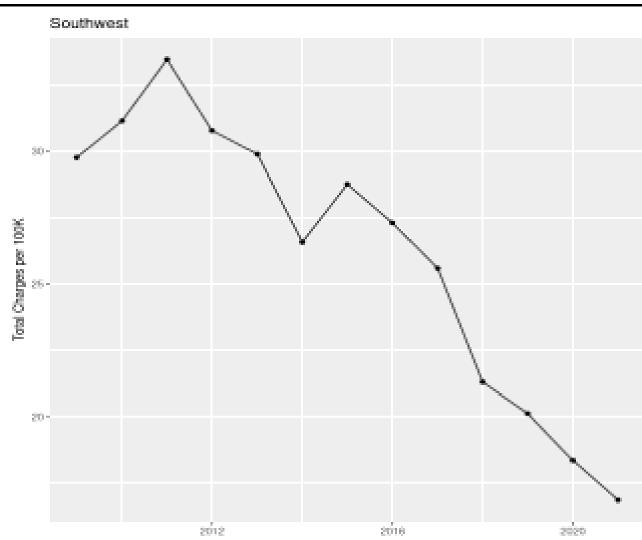



Pacific Coastal
Total Charges per 100K
Year


New England
Total Charges per 100K
Year


The Rocky Mountains
Total Charges per 100K
Year


Southeast
Total Charges per 100K
Year


Southwest
Total Charges per 100K
Year

Output 1. Model 1 Regression Summary

```
Call:
lm(formula = Total_Charges ~ Year + State + Gov_Party + State_Population +
    White_Male_Pop + White_Female_Pop + Black_Male_Pop + Black_Female_Pop +
    Native_American_Male_Pop + Native_American_Female_Pop + Asian_Male_Pop +
    Asian_Female_Pop + Hispanic_Male_Pop + Hispanic_Female_Pop, data)
Residuals:
    Min      1Q  Median      3Q     Max
-970.66  -90.49   -4.90   98.00  921.60
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 5.357e+04  9.068e+03   5.907 6.55e-09 ***
Year                       -2.540e+01  4.591e+00  -5.534 5.13e-08 ***
StateAlaska                -3.029e+03  1.049e+03  -2.887 0.004067 **
.
.
.
StateOregon                -3.630e+03  6.080e+02  -5.971 4.56e-09 ***
StatePennsylvania          -2.250e+03  1.238e+03  -1.817 0.069787 .
StateRhode Island          -2.522e+03  5.485e+02  -4.598 5.46e-06 ***
StateSouth Carolina        -1.528e+03  1.239e+02 -12.336  < 2e-16 ***
StateSouth Dakota          -2.418e+03  7.001e+02  -3.454 0.000601 ***
StateTennessee             -1.553e+03  3.710e+02  -4.186 3.38e-05 ***
StateTexas                  7.886e+03  1.434e+03   5.497 6.23e-08 ***
StateUtah                  -4.033e+03  5.162e+02  -7.813 3.47e-14 ***
StateVermont               -2.795e+03  5.658e+02  -4.940 1.08e-06 ***
StateVirginia              -1.109e+03  5.043e+02  -2.199 0.028346 *
StateWashington            -4.421e+03  7.513e+02  -5.885 7.40e-09 ***
StateWest Virginia         -3.504e+03  4.750e+02  -7.377 7.06e-13 ***
StateWisconsin             -4.575e+03  5.291e+02  -8.648  < 2e-16 ***
StateWyoming               -2.837e+03  5.837e+02  -4.860 1.59e-06 ***
Gov_PartyI                 -6.687e+01  1.366e+02  -0.490 0.624621
Gov_PartyR                 -4.142e+01  2.993e+01  -1.384 0.167070
State_Population            3.115e-04  1.385e-04   2.249 0.024959 *
White_Male_Pop              1.169e-02  3.152e-03   3.708 0.000233 ***
White_Female_Pop           -1.033e-02  3.033e-03  -3.404 0.000719 ***
Black_Male_Pop             -9.279e-03  7.003e-03  -1.325 0.185791
Black_Female_Pop            4.764e-03  6.058e-03   0.786 0.432026
Native_American_Male_Pop    2.955e-01  8.297e-02   3.561 0.000405 ***
Native_American_Female_Pop -3.002e-01  7.290e-02  -4.119 4.48e-05 ***
Asian_Male_Pop             -3.118e-02  8.462e-03  -3.685 0.000255 ***
Asian_Female_Pop            2.645e-02  8.194e-03   3.228 0.001332 **
Hispanic_Male_Pop           1.981e-02  4.637e-03   4.272 2.33e-05 ***
Hispanic_Female_Pop        -2.157e-02  4.617e-03  -4.672 3.86e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 215.5 on 486 degrees of freedom
Multiple R-squared:  0.9897,    Adjusted R-squared:  0.9883
F-statistic: 737.9 on 63 and 486 DF,  p-value: < 2.2e-16
```

(Note: Showing only 15 of 49 state indicator variables)

Output 2. Model 2 Regression Summary

```
Call:
lm(formula = Total_Charges ~ Year + State + State_Population +
    State_Per_Capita_Income + White_Male_Pop + White_Female_Pop +
    Black_Male_Pop + Black_Female_Pop + Native_American_Male_Pop +
    Native_American_Female_Pop + Asian_Male_Pop + Asian_Female_Pop +
    Hispanic_Male_Pop + Hispanic_Female_Pop, data)
Residuals:
    Min      1Q  Median      3Q     Max
-965.40  -89.85   -0.99   99.25  936.57
```

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 4.046e+04  2.146e+04   1.886 0.059925 .  
Year                       -1.882e+01  1.078e+01  -1.745 0.081669 .  
StateAlaska                -2.810e+03  1.094e+03  -2.568 0.010538 *  
.
.
.
StateNew York               6.753e+03  2.405e+03   2.808 0.005187 ** 
StateNorth Carolina         4.417e+03  1.106e+03   3.995 7.48e-05 ***
StateNorth Dakota          -3.245e+03  6.066e+02  -5.349 1.36e-07 ***
StateOregon                -3.584e+03  6.131e+02  -5.847 9.20e-09 ***
StateRhode Island          -2.459e+03  5.584e+02  -4.404 1.31e-05 ***
StateTennessee             -1.576e+03  3.727e+02  -4.227 2.83e-05 ***
StateTexas                  7.856e+03  1.435e+03   5.475 7.03e-08 ***
StateUtah                  -4.044e+03  5.164e+02  -7.831 3.06e-14 ***
StateVermont               -2.746e+03  5.756e+02  -4.770 2.44e-06 ***
StateVirginia              -1.100e+03  5.042e+02  -2.181 0.029674 *  
StateWashington            -4.321e+03  7.514e+02  -5.751 1.57e-08 ***
StateWest Virginia         -3.526e+03  4.750e+02  -7.423 5.14e-13 ***
StateWisconsin             -4.553e+03  5.291e+02  -8.605  < 2e-16 ***
StateWyoming               -2.753e+03  6.081e+02  -4.527 7.52e-06 ***
State_Population            3.004e-04  1.384e-04   2.171 0.030433 *  
State_Per_Capita_Income    -5.045e-03  6.922e-03  -0.729 0.466497    
White_Male_Pop              1.171e-02  3.151e-03   3.715 0.000227 ***
White_Female_Pop           -1.030e-02  3.031e-03  -3.397 0.000738 ***
Black_Male_Pop             -8.041e-03  7.096e-03  -1.133 0.257672    
Black_Female_Pop            3.672e-03  6.144e-03   0.598 0.550319    
Native_American_Male_Pop    2.823e-01  8.269e-02   3.414 0.000694 ***
Native_American_Female_Pop -2.897e-01  7.256e-02  -3.993 7.54e-05 ***
Asian_Male_Pop             -3.086e-02  8.462e-03  -3.647 0.000294 ***
Asian_Female_Pop            2.626e-02  8.195e-03   3.205 0.001439 ** 
Hispanic_Male_Pop           1.932e-02  4.629e-03   4.175 3.54e-05 ***
Hispanic_Female_Pop        -2.108e-02  4.603e-03  -4.580 5.92e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 215.6 on 487 degrees of freedom
Multiple R-squared:  0.9896,	Adjusted R-squared:  0.9883 
F-statistic: 749.2 on 62 and 487 DF,  p-value: < 2.2e-16
```

(Note: Showing only 15 of 49 state indicator variables)

Output 3. Model 3 Regression Summary

```
Call:
lm(formula = Total_Charges ~ Year + State + State_Population + 
    White_Male_Pop + White_Female_Pop + Black_Male_Pop + Black_Female_Pop + 
    Native_American_Male_Pop + Native_American_Female_Pop + Asian_Male_Pop + 
    Asian_Female_Pop + Hispanic_Male_Pop + Hispanic_Female_Pop, data)
Residuals:
    Min      1Q  Median      3Q     Max 
-965.77  -90.01   -2.27   97.66  940.06 
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 5.464e+04  9.032e+03   6.050 2.88e-09 ***
Year                       -2.593e+01  4.573e+00  -5.670 2.44e-08 ***
StateAlaska                -3.045e+03  1.045e+03  -2.914 0.003729 ** 
.
.
.
StateNew York               6.604e+03  2.395e+03   2.757 0.006047 ** 
StateNorth Carolina         4.323e+03  1.098e+03   3.939 9.39e-05 ***
StateNorth Dakota          -3.351e+03  5.883e+02  -5.696 2.12e-08 ***
StateRhode Island          -2.535e+03  5.483e+02  -4.624 4.83e-06 ***
StateSouth Carolina        -1.532e+03  1.235e+02 -12.397  < 2e-16 ***
StateSouth Dakota          -2.446e+03  6.960e+02  -3.514 0.000482 ***
```

```
StateTennessee             -1.550e+03  3.709e+02  -4.179 3.47e-05 ***
StateUtah                  -4.066e+03  5.153e+02  -7.890 2.00e-14 ***
StateVermont               -2.824e+03  5.652e+02  -4.996 8.17e-07 ***
StateVirginia              -1.076e+03  5.030e+02  -2.140 0.032854 *
StateWashington            -4.371e+03  7.479e+02  -5.844 9.33e-09 ***
StateWest Virginia         -3.524e+03  4.747e+02  -7.423 5.13e-13 ***
StateWisconsin             -4.571e+03  5.283e+02  -8.652  < 2e-16 ***
StateWyoming               -2.880e+03  5.824e+02  -4.945 1.05e-06 ***
State_Population            2.892e-04  1.375e-04   2.104 0.035901 *
White_Male_Pop              1.165e-02  3.148e-03   3.699 0.000241 ***
White_Female_Pop           -1.025e-02  3.029e-03  -3.384 0.000771 ***
Black_Male_Pop             -8.888e-03  6.997e-03  -1.270 0.204609
Black_Female_Pop            4.427e-03  6.053e-03   0.731 0.464854
Native_American_Male_Pop    2.922e-01  8.155e-02   3.583 0.000374 ***
Native_American_Female_Pop -2.976e-01  7.172e-02  -4.150 3.93e-05 ***
Asian_Male_Pop             -3.095e-02  8.457e-03  -3.660 0.000280 ***
Asian_Female_Pop            2.628e-02  8.191e-03   3.208 0.001425 **
Hispanic_Male_Pop           1.940e-02  4.625e-03   4.195 3.24e-05 ***
Hispanic_Female_Pop        -2.110e-02  4.601e-03  -4.585 5.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 215.5 on 488 degrees of freedom
Multiple R-squared:  0.9896,    Adjusted R-squared:  0.9883
F-statistic: 762.2 on 61 and 488 DF,  p-value: < 2.2e-16
```

(Note: Showing only 15 of 49 state indicator variables)