UNFCCC Econometric Analysis

United States of America

Paul O'Connor

Executive Summary

Research Purpose

This report's main objective is to investigate the engagement patterns of cities in the United States of America with the United Nations Framework Convention on Climate Change (UNFCCC). By conducting an econometric analysis, it aims to identify the determinants of cities to 1) sign up to the UNFCCC and 2) undertake actions recorded by the UNFCCC. Insights from this analysis can inform policy decision-making, help cities benchmark their climate engagement and act as a reference for similar analyses.

Research Design

The United States of America was chosen as the country of focus. The US' active participation in the UNFCCC, the abundance of high-quality data available, and the interesting political landscape with climate change being a divisive issue facilitate a thorough analysis. The potential determinants of UNFCCC engagement cover demographic, economic, climate change and political features. The analysis leverages statistical techniques such as logistic regression and random forest classification to assess significance, feature importance and explanatory power.

Research Limitations

During this research, several issues limited the effectiveness and validity of the analysis conducted. The number of cities in the US signed up to the UNFCCC meant the sample size for several models was small, which potentially caused problems surrounding statistical testing and model inference. Despite the large amount of data available, when the desired level of granularity was not available county-level variables were used in place of city-level data. This analysis attempted to use a comprehensive set of variables that could explain the variation in UNFCCC engagement, however, there is the possibility of omitted-variable bias leading to distorted results.

Findings

Research question 1 concerned the determinants of cities to sign up for the UNFCCC. Demographic features saw that population had a significantly positive impact and the highest feature importance, median age and education level had significantly negative effects, and ethnic diversity not having any significant influence but having the second highest feature importance. In terms of climate change variables, all had bottom four feature importances, emissions per capita had a significantly negative effect, the number of natural disasters had a significantly negative effect, and the land temperature trend was insignificant. The economic indicators unemployment rate and poverty proportion had significantly negative and positive effects respectively, and median household income was insignificant. Political affiliation significantly influenced signing up to the UNFCCC but ranked lowest in feature importance.

Research question 2 explored the determinants of cities to undertake actions tracked by the UNFCCC. Once again in the demographic features, population was largely positively significant and highly ranked in feature importance, median age and education level were also significant, and ethnic diversity was insignificant but had high feature importance. The climate change variables were largely insignificant and had low feature importance. Out of the economic indicators median household income was positively significant in most models and had the highest feature importance. The presence of a Republican majority was insignificant in most models and had the lowest feature importance in all random forests.

Research Implications

In research question 1, it was concluded that cities with larger, younger and more educated populations are more likely to sign up for the UNFCCC. The same can be said for cities more exposed to the immediate effects of climate change. The economic prosperity of cities may influence UNFCCC engagement, given the negative relationship between the unemployment rate and signing up for the framework. The significance of a Republican majority illustrated that the partisan divide in the perception of climate change is also present in city UNFCCC engagement.

Population exhibited consistent explanatory power across the actions examined in research question 2, with larger, younger, more educated cities being more likely to undertake UNFCCC actions. The climate change features had minimal influence, which is potentially due to the state-level granularity of those variables. Unlike in research question 1, cities with higher median household income were significantly more likely to engage with the UNFCCC. The political affliction of a city proved to be ineffective in explaining the nuances of the individual actions recorded by the UNFCCC.

This report can hold significance for the UNFCCC, highlighting areas of focus for improving engagement among cities unlikely to participate. The findings also point to the need for further research and the possibility of exploring more variables and countries.

Research Purpose

The United Nations Framework Convention on Climate Change (UNFCCC) is an international treaty first opened for signature in 1992. The main objective of the UNFCCC is the "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system" (United Nations, 1992). Since the treaty's inception 34,648 actors, including companies, investors, organizations, regions, cities, and countries, have signed up to engage in climate action, mitigating the negative human effect on our planet's ecosystem (Unfccc.int, 2022). The UNFCCC commitments include periodically publishing national inventories of human-caused emissions and programmes containing measures to mitigate climate change, promoting the development of technologies to curb emissions and public awareness of climate change, and accounting for climate change in social, economic, and environmental policies and actions.

The focus of this report will be on the cities in the United States of America's engagement with the UNFCCC, with the objective of conducting a comprehensive econometric analysis of the determinants of US cities

1. to sign up to the UNFCCC

and, of those that are signed up,

2. to undertake individual actions recorded by the UNFCCC.

The results of this analysis hold significance for stakeholders, including the UNFCCC, cities, and other countries. The UNFCCC can benefit from insights into the determinants of city engagement by informing policymaking and efforts to increase participation. City administrations looking to benchmark their engagement can do so using the key determinants of engagement to find similar counterparties. Countries other than the US looking to conduct a similar analysis of their cities' engagement with the UNFCCC can use the findings of this report to shape their research.

Research Design

Country of Focus Selection

The United States of America was selected as the focus of this analysis due to several factors including the four below.

1. UNFCC Participation
   The US has 298 cities, city councils, municipal organisations, and local authorities, representing a total population of 88.3 million people, which are actors in the UNFCCC. The US has the fifth most cities signed up for the UNFCCC of any country in the world and the most of any English-speaking country. The high level of climate engagement among US cities is a notable reason for the country's inclusion in this econometric analysis.

2. Data Availability
   To conduct a comprehensive analysis, a broad range of data is required. The US has a large range of publicly available data from government sources such as the US Bureau of Labor Statistics, the U.S. Department of Commerce, and the US Census Bureau. Therefore, given the vast amount of easily accessible data, the US is an obvious choice for this analysis.

3. US Politics and Climate Change
   The political landscape, especially surrounding the issue of climate change, presents a compelling area of focus for this analysis. Climate change is a topical issue in the US with recent legislation, such as the Inflation Reduction Act of 2022, promoting major investment in clean energy. 78% of Democrats view climate change as a "major threat" compared to 23% of Republicans, highlighting the partisan divide in opinions on the issue (Tyson, Funk and Kennedy, 2023). This difference has been increasing over the past ten years, with the republican proportion increasing one percentage point in the period compared to the twenty percentage point increase in the Democratic figure (Fagan and Huang, 2019).

4. Demographics
   With a diverse population of 333.3 million encompassing ranges of ethnicities, ages, occupations, political views, and over 15,000 cities with more than 1,000 inhabitants, the US offers scope for a comprehensive analysis.

Data / Variables

A range of indicators were selected for inclusion in this econometric analysis due to their potential discriminatory power in explaining the engagement patterns of cities with the UNFCCC.

1. Demographic Features
   Across many disciplines demographic characteristics are common differentiators of countries, regions, and cities. The motivation for the inclusion of the factors below is that a younger, more diverse, and more educated population may show a higher level of support for climate engagement. A 2021 survey of 10,000 people aged 16-25 from 10 countries found 59% of young people were "extremely worried" about climate change and that percentage increased to 84% for respondents who were "at least moderately worried" (Hickman et al., 2021). Education levels can have a statistically significant impact on perceptions of climate change, with 56% of Americans with a high school education or less thinking climate change is a major threat to their country while the percentage of their more highly educated compatriots is 62% (Fagan and Huang, 2019). Younger adults in the US are also more open to eliminating fossil fuels (Tyson, Funk and Kennedy, 2023).
   - Population
   - Median Age
   - Ethnic Diversity
   - Education Level

2. Climate Change Features
   Given that climate change is at the centre of this analysis, the reason for including variables related to the issue is that cities exposed to climate change may be more likely to engage with the UNFCCC. In recognition of the immediate impact of climate change a city faces, it may acknowledge the need for a collective effort to

mitigate those effects. Perception of these effects does differ by region, with 51% of Americans from the Pacific region, where wildfires and earthquakes occur often, saying that they see the local impact of climate change compared to 38% in the rest of the country (Tyson, Funk and Kennedy, 2023). The variables in this category are as follows.

- Emissions Per Capita
- Natural Disasters
- Land Temperature Trends

3. Economic Features

The rationale behind this analysis incorporating economic variables is that the prosperity of a city may influence its participation in the UNFCCC. That is that cities facing more immediate issues may have fewer resources to focus on climate change engagement. Below are the three economic variables selected for inclusion in this analysis.

- Unemployment Rate
- Median Household Income
- Proportion of Population in Poverty

4. Political Feature

As mentioned as motivation for the selection of the US as the focus of this analysis, political affiliation has an impact on a citizen's opinions on climate change. The reason for the inclusion of political data is that a city with more democratic voters may be more likely to engage with UNFCCC. Please see the political variable included in the analysis below.

- 2020 Presidential Election Results

A full detailed list of the data sources and the corresponding variables associated with them can be found in the appendix.

Methodology

The analysis leverages several statistical techniques to assess the deterministic power of the variables of interest including:

1. Logistic Regression – statistical significance
2. Forward Selection – variables selected and their order
3. Random Forest Classification – feature importance

Detailed descriptions of all statistical techniques used in the analysis can be found in the Methodology section of the Appendix.

The analysis was performed using Python and leveraged the functionality of various packages including Pandas, NumPy, Matplotlib, Seaborn, Statsmodels, Scikit-Learn and GeoPandas. A public GitHub repository was used to maintain version control and host the analysis codebase.

Research Limitations

Several issues, encountered during the completion of this project, affected the validity and accuracy of this analysis. These limitations should be considered when evaluating the findings of this report.

1. Sample Size

   Despite the US having a high number of cities among the UNFCCC actors list relative to other countries, the sample size of the analysis was still small. Only 202 UNFCCC cities were used in the analysis after joining and data processing. This created a clear data imbalance when analysing the first research question, with UNFCCC cities only accounting for only 1.3% of the data set. The small number of UNFCCC cities limited the analysis of the second research question where the entire sample size of the models was 202 observations. While smaller sample sizes have minimal impact on descriptive statistics, if a sample is not sufficiently large it can have significant effects on results of statistical inference, for example, tests of coefficient significance in the logistic regression models that are used in this analysis (Chandrasekharan et al., 2019).

2. Data Availability

   Once again, while the US has a relatively wide range of data compared to other countries, there were limitations that affected the analysis.

   The lack of data at a city-level was an inherent challenge during this analysis. Most government data sources are mostly at less granular levels, such as country, region, state and county levels. While county-level data is a reasonable proxy and was used wherever available, it does not perfectly capture the characteristics of all cities. Certain counties encompass multiple cities, others only cover subsections of larger metropolitan areas, and several cities span across multiple counties or states, all of which cause mapping issues. Most counties originate from the late 19^{th} century with little to no change since then and therefore may not capture modern population dynamics. The number of counties in the US today is 3,142 compared to 3,041 in 1920. Counties' objectives, responsibilities and sizes differed across different regions of the US. For example, more compact settlements in the northeast meant counties encompassed multiple towns and cities while the more dispersed settlements of the south meant counties traditionally only contained single towns (NACo, 2012).

   The lack of up-to-date data was also a challenge during this analysis. For example, the land temperature dataset only contained US state data up to 2013. Similarly, 2020 presidential election data was used instead of potential recent polling data for the upcoming 2024 election. These older datasets limit the potential for recent trends to be captured within the analysis.

3. Omitted Variable Bias

   While an effort has been made to cover a broad spectrum of potential key determinants of UNFCCC engagement, there is the possibility that some variables that could have a significant impact on the dependent variables in this analysis. Not including useful variables can lead to models attributing their effect to variables that are included (Wilms et al., 2021).

Findings

Detailed findings of both research questions and all associated models can be found in the Results section of the Appendix. A summary of the findings is presented below.

Research Question 1

The first research question concerns the determinants of a city to sign up for the UNFCCC.

1. Demographic Features
   - Population was significant at the 1% level in both the full and the forward selection logistic regression models. In both cases it had a positive effect on the log odds of the dependent variable, meaning that as its population increases, the probability of a city being signed up to the UNFCCC increases. It was the first variable selected in the forward selection logistic regression model. It was ranked as the variable with the highest feature importance in the random forest classification model.

   - Median Age was significant at the 5% level in both the full and the forward selection logistic regression models. In both cases it had a negative effect on the log odds of the dependent variable, meaning that as its median age increases, the probability of a city being signed up to the UNFCCC decreases. It was ranked as the variable with the fourth highest feature importance in the random forest classification model.

   - Ethnic Diversity was insignificant in the full logistic regression model and not selected in the forward selection models. It was ranked as the variable with the second highest feature importance in the random forest classification model.

   - Education Level was significant at the 1% level in both the full and the forward selection logistic regression models. In both cases it had a negative effect on the log odds of the dependent variable, meaning that as the proportion of its population without less than a high school diploma increases, the probability of a city being signed up to the UNFCCC decreases. It was ranked as the variable with the third highest feature importance in the random forest classification model.

2. Climate Change Features
   - Emissions Per Capita was significant at the 1% level in both the full and the forward selection logistic regression models. In both cases it had a negative effect on the log odds of the dependent variable, meaning that as its average emissions per capita increases, the probability of a city being signed up to the UNFCCC decreases. It was ranked as the variable with the eighth highest feature importance in the random forest classification model.

   - Natural Disasters was significant at the 1% level in both the full and the forward selection logistic regression models. In both cases it had a positive effect on the log odds of the dependent variable, meaning that as the number of state-level natural disasters increases, the probability of a city being signed up to the UNFCCC increases. It was ranked as the variable with the second lowest feature importance in the random forest classification model.

   - The Land Temperature Trends was insignificant in the full logistic regression model and not selected in the forward selection models. It was ranked as the variable with the third lowest feature importance in the random forest classification model.

3. Economic Features
   - The Unemployment Rate was significant at the 5% level in both the full and the forward selection logistic regression models. In both cases it had a negative effect on the log odds of the dependent variable, meaning that as its unemployment rate increases, the probability of a city being signed up to the UNFCCC decreases. It was ranked as the variable with the seventh highest feature importance in the random forest classification model.

   - Median Household Income was insignificant in the full logistic regression model and not selected in the forward selection models. It was ranked as the variable with the fifth highest feature importance in the random forest classification model.

- Proportion of Population in Poverty was significant at the 10% level in the forward selection logistic regression model but not the full model. It had a positive effect on the log odds of the dependent variable, meaning that as the proportion of its population that is in poverty increases, the probability of a city being signed up to the UNFCCC increases. The inclusion of additional variables in the full model possibly attenuated its explanatory power. It was ranked as the variable with the sixth highest feature importance in the random forest classification model.

4. Political Feature
   - 2020 Presidential Election Results were significant at the 1% level in both the full and the forward selection logistic regression models. In both cases it had a negative effect on the log odds of the dependent variable, meaning if a Republican majority is present, the probability of a city being signed up to the UNFCCC decreases. It was ranked as the variable with the lowest feature importance in the random forest classification model.

Presented below are the full results of the models used in research question 1.

Dependent Variable	**UNFCCC**	
	Full	**Forward Selection**
constant	-0.3804	-0.2816
population	0.0000***	0.0000***
redCounty	-1.6097***	-1.6067***
unemploymentRate	-27.5375**	-26.5005**
povertyProp	4.8705	4.6266*
tempDiff	-0.0289	
numDisasters	0.0036***	0.0037***
avgEmissionsPerCapita	-0.0652***	-0.0653***
lessThanHighSchoolProp	-11.8898***	-12.1238***
medianHouseholdIncome	0.0000	
medianAge	-0.0449**	-0.0461**
whiteProp	0.0399	
AIC	1511	**1508**
BIC	1587	**1569**

* *p<0.05, ** p<0.01, *** p<0.001*

Feature	**UNFCCC**	
	Importance	**Rank**
population	0.633	1
whiteProp	0.056	2
lessThanHighSchoolProp	0.052	3
medianAge	0.049	4
medianHouseholdIncome	0.046	5
povertyProp	0.041	6
unemploymentRate	0.037	7
avgEmissionsPerCapita	0.029	8
tempDiff	0.024	9
numDisasters	0.023	10
redCounty	0.010	11

Research Question 2

The second research question concerns the determinants of a city to undertake individual actions recorded by the UNFCCC. There were ten different individual actions recorded by the UNFCCC across three categories: Engagements, Climate Focus Actions and Progress Tracking.

1. Demographic Features
   - Population was significant at the 1% level in nine of the twenty logistic regression models. In all models in which it was significant, it had a positive effect on the log odds of the dependent variable, meaning that as its population increases, the probability of a city undertaking those actions recorded by the UNFCCC increases. It was the first additional variable selected in eight of the ten forward selection logistic regression models. It was ranked as the variable with the highest feature importance in all ten random forest classification models.

   - Median Age was significant at the 5% level in three and at the 10% level in two logistic regression models. In all models in which it was significant, it had a negative effect on the log odds of the dependent variable, meaning that as its median age increases, the probability of a city undertaking those actions recorded by the UNFCCC decreases. It was selected in three forward selection models and dropped its significance level in the full model for all but one of those (Mitigations). In terms of the random forest classification models, it had an average feature importance ranking of 3.3, its lowest ranking was fifth (Initiative Participations) and its highest ranking was second (Commitments and Mitigations).

   - Ethnic Diversity was insignificant in all the models and was therefore not included in any forward selection models. This means that changes in the unemployment rate have no effect on the log odds of the dependent variables. In terms of the random forest classification models, it had an average feature importance ranking of fourth, its lowest ranking was seventh (Impact) and its highest ranking was second (Initiative Participation and Finance Actions).

   - Education Level was significant at the 5% level in four and at the 10% level in four logistic regression models. In all models in which it was significant, it had a negative effect on the log odds of the dependent variable, meaning that as the proportion of its population without less than a high school diploma increases, the probability of a city undertaking those actions recorded by the UNFCCC decreases. It was selected in four forward selection models and dropped its significance level in the full model for all but two of those (Emissions Inventory and Impact). In terms of the random forest classification models, it had an average feature importance ranking of 5.2, its lowest ranking was seventh (Impact, Actions Undertaken, Initiative Participation, Mitigations, Adaptations and Climate Action Plans) and its highest ranking was second (Emissions Inventory).

2. Climate Change Features
   - Emissions Per Capita was significant at the 5% level in one logistic regression model (Actions Undertaken). In the model in which it was significant, it had a positive effect on the log odds of the dependent variable, meaning that as the average emissions per capita increase, the probability of a city undertaking those actions recorded by the UNFCCC increases. It was selected in two forward selection models, one in which it dropped its significance in the full model (Actions Undertaken) and one in which it was not significant (Adaptations). In terms of the random forest classification models, it had an average feature importance ranking of 8.7, its lowest ranking was tenth (Impact and Finance Actions) and its highest ranking was eighth (Commitments, Emissions Inventory, Actions Undertaken, Initiative Participation, Adaptations and Climate Action Plans).

   - Natural Disasters was significant at the 5% level in two logistic regression models and at the 1% level in one model (Mitigations). In the models in which it was significant, it had opposite effects on the log odds of the dependent variable (positive for Impact and negative for Mitigations). It was selected in two forward selection models and dropped its significance level in the full model for all but two of those (Impact). In terms of the random forest classification models, it had an average feature importance ranking of 9.4, its lowest ranking was tenth (including Impact and Mitigations) and its highest ranking was ninth.

   - Land Temperature Trend was significant at the 10% level in one logistic regression model (Actions Undertaken). In the model in which it was significant, it had a positive effect on the log odds of the dependent variable, meaning that as the average July-January temperature range difference between 2000-2013 and 1940-1960 increases, the probability of a city undertaking actions recorded by the UNFCCC increases. It was not included in any forward selection models. In terms of the random forest classification models, it had an average feature importance ranking of eighth, its lowest ranking was tenth (Actions Undertaken, Emissions Inventory and Climate Action Plans) and its highest ranking was fourth (Initiative Participation).

3. Economic Features

- Unemployment Rate was insignificant in all the models and was therefore not included in any forward selection models. This means that changes in the unemployment rate have no effect on the log odds of the dependent variables. In terms of the random forest classification models, it had an average feature importance ranking of 6.4, its lowest ranking was eighth (Risk Assessments) and its highest ranking was third (Finance Actions).

- Median Household Income was significant at the 1% level in six logistic regression models. In all models in which it was significant, it had a positive effect on the log odds of the dependent variable, meaning that as its median household income increases, the probability of a city undertaking those actions recorded by the UNFCCC increases. It was selected in six forward selection models and dropped its significance level in the full model for all but three of those (Emissions Inventory, Risk Assessments and Climate Action Plans). In terms of the random forest classification models, it had an average feature importance ranking of 3.3, its lowest ranking was seventh (Finance Actions) and its highest ranking was second (Actions Undertaken, Adaptations, Risk Assessments and Climate Action Plans).

- Proportion of Population in Poverty was significant at the level 1% in three logistic regression models (Actions Undertaken and Emissions Inventory) and at the 5% and 10% levels in one model each. In the models in which it was significant, it had opposite effects on the log odds of the dependent variable (positive for Emissions Inventory and Initiative Participation and negative for Actions Undertaken). It was selected in two forward selection models but dropped two significance levels in one of the full models (Actions Undertaken). In terms of the random forest classification models, it had an average feature importance ranking of 5.7, its lowest ranking was ninth (Initiative Participation) and its highest ranking was second (Impact).

4. Political Feature

- 2020 Presidential Election Results were significant at the 5% and 10% levels in one model each. In all models in which it was significant, it had a negative effect on the log odds of the dependent variable, meaning that if a Republican majority is present, the probability of a city undertaking those actions recorded by the UNFCCC decreases. It was selected for one forward selection model but its significance dropped in the full model (Initiative Participation). It was ranked as the variable with the lowest feature importance in all ten random forest classification models.

Presented below are the full results of the models used in research question 2.

Dependent Variable	hasCommitments		hasActionsUndertaken		hasInitiativeParticipations		hasClimateActionPlans	
	Full	**Forward Selection**	**Full**	**Forward Selection**	**Full**	**Forward Selection**	**Full**	**Forward Selection**
constant	7.1363	0.2236	2.4800	7.7181***	-7.0407	1.5515***	-0.6963	2.9802
population	0.0000***	0.0000***	0.0000***	0.0000***	0.0000**	0.0000***	0.0000*	0.0000*
redCounty	-0.2276		0.2218		-0.7139	-0.9902**	-0.0495	
unemploymentRate	6.4997		22.8621		-45.8746		22.0077	
povertyProp	-6.9890		-21.0028*	-32.2155***	29.7740**		6.0801	
tempDiff	0.1049		0.4889*		-0.3438		0.1765	
numDisasters	-0.0018		-0.0038		0.0003		-0.0015	
avgEmissionsPerCapita	-0.0416		0.0677	0.0917**	0.0123		-0.0034	
lessThanHighSchoolProp	-10.9615*	-10.1646**	-2.8520		-0.5495		-8.9975	-8.4656**
medianHouseholdIncome	0.0000	0.0000	0.0000		0.0000**		0.0000**	0.0000**
medianAge	-0.0938		-0.1402*	-0.1155**	0.0598		-0.0949	-0.0839*
whiteProp	-1.1494		2.2028		0.6053		1.8492	
AIC	217	**206**	157	**149**	186	**178**	243	**231**
BIC	256	**219**	196	**165**	226	**188**	282	**248**

** p<0.05, ** p<0.01, *** p<0.001*

Dependent Variable	hasMitigations		hasAdaptations		hasFinanceActions	
	Full	**Forward Selection**	**Full**	**Forward Selection**	**Full**	**Forward Selection**
constant	1.5733	2.4879	0.7971	-3.1342***	5.3011	-2.7623***
population	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
redCounty	0.6430		-0.7186*		0.4428	
unemploymentRate	29.5499		1.2851		78.4613	
povertyProp	-2.4839		-7.5449		-7.0380	
tempDiff	0.3517		0.1111		0.3943	
numDisasters	-0.0033	-0.0070***	-0.0034		0.0069**	
avgEmissionsPerCapita	0.0442		0.0450	0.0573	-0.0603	
lessThanHighSchoolProp	-7.7900		1.3696		-28.9080*	
medianHouseholdIncome	0.0001*	0.0000***	0.0000	0.0000***	0.0000	
medianAge	-0.1556**	-0.1128**	-0.0748		-0.0901	
whiteProp	1.4624		1.0639		-2.7961	
AIC	187	**179**	229	**220**	115	**104**
BIC	226	**195**	268	**233**	154	**111**

** p<0.05, ** p<0.01, *** p<0.001*

Dependent Variable	hasEmissionInventory		hasRiskAssessments		hasImpact	
	Full	**Forward Selection**	**Full**	**Forward Selection**	**Full**	**Forward Selection**
constant	-8.3588**	-6.5556***	-3.8236	-2.7077***	-9.0717	-4.7994**
population	0.0000***	0.0000***	0.0000***	0.0000***	0.0000	
redCounty	-0.3825		-0.2678		0.1913	
unemploymentRate	20.3213		28.2781		9.5872	
povertyProp	22.8593***	20.8728***	6.9209		24.4481	19.4192
tempDiff	0.2441		0.3567		-0.0505	
numDisasters	-0.0006		-0.0018		0.0121**	0.0132**
avgEmissionsPerCapita	0.0134		0.0004		0.0069	
lessThanHighSchoolProp	-12.0784**	-11.1956**	-4.7900		-20.5532*	-25.3963*
medianHouseholdIncome	0.0001***	0.0001***	0.0001***	0.0000***	0.0000	
medianAge	-0.0315		-0.0718		-0.0529	
whiteProp	1.6235		1.2963		3.8310	
AIC	242	**231**	256	**245**	92	**78**
BIC	281	**247**	296	**255**	131	**91**

** p<0.05, ** p<0.01, *** p<0.001*

Feature	hasCommitments		hasActionsUndertaken		hasEmissionInventory		hasInitiativeParticipations		hasImpact	
	Importance	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**
population	0.244	1	0.250	1	0.278	1	0.223	1	0.267	1
redCounty	0.010	11	0.005	11	0.012	11	0.030	11	0.006	11
unemploymentRate	0.080	6	0.062	7	0.069	7	0.083	7	0.092	5
povertyProp	0.079	7	0.102	4	0.077	6	0.079	9	0.132	2
tempDiff	0.059	9	0.042	10	0.048	10	0.084	4	0.049	8
numDisasters	0.053	10	0.043	9	0.050	9	0.059	10	0.048	9
avgEmissionsPerCapita	0.065	8	0.054	8	0.053	8	0.082	8	0.046	10
lessThanHighSchoolProp	0.094	4	0.077	6	0.108	2	0.083	6	0.087	6
medianHouseholdIncome	0.105	3	0.151	2	0.100	5	0.088	3	0.095	4
medianAge	0.129	2	0.125	3	0.103	3	0.084	5	0.103	3
whiteProp	0.080	5	0.091	5	0.102	4	0.105	2	0.077	7

Feature	hasMitigations		hasAdaptations		hasRiskAssessments		hasClimateActionPlans		hasFinanceActions	
	Importance	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**	**Importance**	**Rank**
population	0.233	1	0.280	1	0.247	1	0.293	1	0.246	1
redCounty	0.008	11	0.014	11	0.009	11	0.014	11	0.015	11
unemploymentRate	0.061	7	0.063	7	0.074	8	0.072	7	0.100	3
povertyProp	0.083	5	0.091	4	0.078	7	0.078	5	0.072	8
tempDiff	0.059	8	0.057	9	0.080	6	0.058	10	0.087	6
numDisasters	0.051	10	0.048	10	0.065	9	0.059	9	0.065	9
avgEmissionsPerCapita	0.059	9	0.059	8	0.055	10	0.060	8	0.039	10
lessThanHighSchoolProp	0.081	6	0.071	6	0.081	5	0.076	6	0.096	5
medianHouseholdIncome	0.116	3	0.129	2	0.124	2	0.110	2	0.076	7
medianAge	0.148	2	0.098	3	0.087	4	0.083	4	0.097	4
whiteProp	0.102	4	0.091	5	0.100	3	0.095	3	0.109	2

Research Implications

Research Question 1

The inclusion of demographic features was useful in explaining whether US cities signed up to the UNFCCC. Population emerged as a key determinant, indicating that cities with more inhabitants were more likely to sign up. Median age and education level having negative effects on engagement supports the hypothesis that younger, more educated demographic groups exhibit a heightened focus on climate change. While ethnic diversity was insignificant in the logistic regression model, its high feature importance may indicate its explanatory power. However, with the absence of the direction of the effect, due to the opacity of ensemble methods, it is unclear whether more ethnically diverse cities have a higher propensity to sign up to the UNFCCC.

In terms of the climate change features, emissions per capita and frequency of natural disasters' significant impacts on signing up for the UNFCCC, point to cities that are more exposed to the implications of climate change are more focused on the issue. The land temperature trend not showing a significant impact may suggest that temperature increases are not affecting climate change engagement, however, there could be explanatory power that is not being captured as a result of the variable construction.

The negative influence the unemployment rate exhibited on a city's propensity to sign up for the UNFCCC was in line with the theory that better economic prosperity of a city may increase climate engagement. This was challenged by the poverty proportion having a significantly positive relationship. However, this was only in the forward selection model and the significance was not present in the full model, so may be a result of falsely attributed explanatory power in the model with fewer variables.

Political affiliation proved to be important in determining whether a city was signed up to the UNFCCC. The presence of a Republican majority had a negative effect, demonstrating that the partisan divide in the perception of climate change is also present in city UNFCCC engagement.

Research Question 2

The demographic feature, population was consistently useful in explaining cities' engagement across the majority of actions recorded by the UNFCCC, further illustrating that more populous cities are more active with respect to climate action. Median age and education level had negative effects on some of the actions undertaken, supporting the notion of higher climate advocacy of younger, more educated population influencing city engagement. Ethnic diversity's high feature importance indicates an influence on actions undertaken, however, the lack of direction of the effect limits the conclusions that can be drawn.

The climate change features had relatively low explanatory power in determining cities' UNFCCC actions undertaken. The minimal significance and feature importance point to the variables not being able to explain the intricacies of the different actions which may be due to the lack of granularity of the data sources, all at the state-level.

The economic indicators saw median household income and the poverty proportion influencing several UNFCCC actions. Median household income had a positive effect, indicating that cities with wealthier inhabitants are more likely to undertake certain actions. The unemployment rate being insignificant and the poverty proportion having effects in different directions may suggest that, at the more nuanced level of individual actions, economic prosperity is less effective in explaining climate engagement patterns.

The presence of a Republican majority proved to be less effective in explaining the variation in actions undertaken that are recorded by the UNFCCC. The lowest feature importance and lack of significance in most models show the minimal effect political affiliation has on cities' climate actions.

Overall, these findings hold significance for the UNFCCC and could be used to increase climate engagement in the US by focusing on cities that do not fit the characteristics of those that are likely to engage. For example, advocating for lower population, older, and less educated cities to sign up for and engage with the UNFCCC.

Further analysis of these research questions could build on the findings of this report, possibly broadening the investigation to more counties or variables and examining the underlying factors driving the significant explanatory power of some of the variables.

Appendix

Results

Research Question 1

The first research question concerns the determinants of a city to sign up to the UNFCCC. The sample size for the models in this analysis is 14,546, with 1.35% of the data with a positive value for the dependent variable 'UNFCCC'. This is 'true' when the city is present in the UNFCCC actor tracking.

1. Logistic Regression

   The logistic regression model of all observations had seven statistically significant coefficients, and five at the 1% level. A city's population and number of natural disasters has a positive effect on the log odds of that city being signed up to the UNFCCC. The presence of a Republican majority had a negative effect on the log odds and increases in unemployment rate, emissions per capita, proportion of people with less than a high school diploma and the median age also resulted in decreases in the log odds of being signed up to the UNFCCC. The proportion of a city's population which is in poverty, the land temperature trend, its median household income, and the proportion of the population categorised as white all had statistically insignificant impacts on whether the city was signed up to the UNFCCC.

Dependent Variable	**UNFCCC**	
	Full	**Forward Selection**
constant	-0.3804	-0.2816
population	0.0000***	0.0000***
redCounty	-1.6097***	-1.6067***
unemploymentRate	-27.5375**	-26.5005**
povertyProp	4.8705	4.6266*
tempDiff	-0.0289	
numDisasters	0.0036***	0.0037***
avgEmissionsPerCapita	-0.0652***	-0.0653***
lessThanHighSchoolProp	-11.8898***	-12.1238***
medianHouseholdIncome	0.0000	
medianAge	-0.0449**	-0.0461**
whiteProp	0.0399	
AIC	1511	**1508**
BIC	1587	**1569**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

   The forward selection model included eight statistically significant coefficients, and the same five at the 1% level as the full model. This model offers a better fit with a lower AIC and BIC. The proportion of a city's population which is in poverty gained significance at the 10% level. Introducing additional variables in the full model may attenuate the explanatory power attributed to the poverty proportion in the forward selection model. Population was the first additional variable selected in the model, which possibly indicates its relative importance in explaining UNFCCC engagement compared to the other variables.

3. Random Forest Classification

   The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city signed up to the UNFCCC. The proportion of the population categorised as white and the median household income are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index despite both being significant at the 1% level in the logistic regression models. The land temperature trend was the third least important variable and was insignificant in the logistic regression models indicating its lack of discriminatory power in explaining UNFCCC participation.

	UNFCCC	
Feature	**Importance**	**Rank**
population	0.633	1
whiteProp	0.056	2
lessThanHighSchoolProp	0.052	3
medianAge	0.049	4
medianHouseholdIncome	0.046	5
povertyProp	0.041	6
unemploymentRate	0.037	7
avgEmissionsPerCapita	0.029	8
tempDiff	0.024	9
numDisasters	0.023	10
redCounty	0.010	11

Research Question 2

The second research question concerns the determinants of a city to undertake individual actions recorded by the UNFCCC. The sample size for the models in this analysis is 197. In each case dependent variable is 'true' when the city has undertaken specific actions. The results are split into four categories of actions tracked by the UNFCCC:

a) Engagements - Commitments, Actions Undertaken, Initiative Participation, Climate Action Plans
b) Climate Focus Actions – Mitigation, Adaptation, Finance
c) Progress Tracking – Emissions Inventory, Risk Assessment, Impact

a) Engagements – Commitments

UNFCCC commitments encompass creating and executing local and regional initiatives to reduce greenhouse gas emissions, adjust to the effects of climate change, and integrate climate concerns into pertinent policies and measures. The dependent variable 'hasCommitments' is positive for 81.2% of observations used in the models.

1. <u>Logistic Regression</u>

The logistic regression model of all observations had two statistically significant coefficients, and one at the 1% level. A city's population has a positive effect on the log odds of that city having commitments tracked by the UNFCCC. The proportion of people with less than a high school diploma was significant at the 10% level and resulted in decreases in the log odds of having UNFCCC commitments. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	**hasCommitments**	
	Full	**Forward Selection**
constant	7.1363	0.2236
population	0.0000***	0.0000***
redCounty	-0.2276	
unemploymentRate	6.4997	
povertyProp	-6.9890	
tempDiff	0.1049	
numDisasters	-0.0018	
avgEmissionsPerCapita	-0.0416	
lessThanHighSchoolProp	-10.9615*	-10.1646**
medianHouseholdIncome	0.0000	0.0000
medianAge	-0.0938	
whiteProp	-1.1494	
AIC	217	**206**
BIC	256	**219**

** p<0.05, ** p<0.01, *** p<0.001*

2. <u>Forward Selection</u>

The forward selection model included the same two statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was significant at the 1% level and the first additional variable selected in the model, which possibly indicates its relative importance in explaining UNFCCC commitments compared to the other variables. The proportion of people with less than a high school diploma gained significance to the 5% level potentially due to the inclusion of additional variables in the full model attenuating its explanatory power. The median household income was included in the model, meaning that, despite it being statistically insignificant, it decreases the AIC.

3. <u>Random Forest Classification</u>

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has commitment tracked by the UNFCCC. The proportion of the population categorised as white, the median age and the median household income are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority, the number of natural disasters and the land temperature trend were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasCommitments	
	Importance	**Rank**
population	0.244	1
medianAge	0.129	2
medianHouseholdIncome	0.105	3
lessThanHighSchoolProp	0.094	4
whiteProp	0.080	5
unemploymentRate	0.080	6
povertyProp	0.079	7
avgEmissionsPerCapita	0.065	8
tempDiff	0.059	9
numDisasters	0.053	10
redCounty	0.010	11

a) Engagements – Actions Undertaken

UNFCCC actions describe the undertaking any of the Climate Focus Actions: Adaptation, Mitigation or Finance. The dependent variable 'hasActionsUndertaken' is positive for 90.4% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had four statistically significant coefficients, and one at the 1% level. A city's population has a positive effect on the log odds of that city having actions undertaken that are tracked by the UNFCCC. The proportion of a city's population which is in poverty and the median age were significant at the 10% level and resulted in decreases in the log odds of having UNFCCC actions undertaken. The land temperature trend was also significant at the 10% level but has a positive coefficient. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasActionsUndertaken	
	Full	**Forward Selection**
constant	2.4800	7.7181***
population	0.0000***	0.0000***
redCounty	0.2218	
unemploymentRate	22.8621	
povertyProp	-21.0028*	-32.2155***
tempDiff	0.4889*	
numDisasters	-0.0038	
avgEmissionsPerCapita	0.0677	0.0917**
lessThanHighSchoolProp	-2.8520	
medianHouseholdIncome	0.0000	
medianAge	-0.1402*	-0.1155**
whiteProp	2.2028	
AIC	157	**149**
BIC	196	**165**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included five statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was significant at the 1% level and the first additional variable selected in the model, which possibly indicates its relative importance in explaining UNFCCC actions undertaken compared to the other variables. The constant gained significance at the 1% level indicating a baseline log odds of 7.7181 when all other variables are zero. The proportion of a city's population which is in poverty, the city's emissions per capita and the median age all had increased significance due to the inclusion of additional variables in the full model attenuating their explanatory power.

3. Random Forest Classification

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has actions undertaken that are tracked by the UNFCCC. The proportion of the population categorised as white and the median household income are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority, the number of natural disasters and the land temperature trend were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasActionsUndertaken	
	Importance	Rank
population	0.250	1
medianHouseholdIncome	0.151	2
medianAge	0.125	3
povertyProp	0.102	4
whiteProp	0.091	5
lessThanHighSchoolProp	0.077	6
unemploymentRate	0.062	7
avgEmissionsPerCapita	0.054	8
numDisasters	0.043	9
tempDiff	0.042	10
redCounty	0.005	11

a) Engagements – Initiative Participation

The dependent variable 'hasInitiativeParticipations' is positive for 96.4% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had three statistically significant coefficients, and none at the 1% level. A city's population, the proportion of its population which is in poverty and the median household income are all significant at the 5% level and have positive effects on the log odds of that city participating in UNFCCC initiatives. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasInitiativeParticipations	
	Full	Forward Selection
constant	-7.0407	1.5515***
population	0.0000**	0.0000***
redCounty	-0.7139	-0.9902**
unemploymentRate	-45.8746	
povertyProp	29.7740**	
tempDiff	-0.3438	
numDisasters	0.0003	
avgEmissionsPerCapita	0.0123	
lessThanHighSchoolProp	-0.5495	
medianHouseholdIncome	0.0000**	
medianAge	0.0598	
whiteProp	0.6053	
AIC	186	**178**
BIC	226	**188**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included three statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was significant at the 1% level but not the first additional variable selected in the model. The presence of a Republican majority was the first selected which possibly indicates its relative importance in explaining UNFCCC initiative participation compared to the other variables. The constant gained significance at the 1% level indicating a baseline log odds of 1.5515 when all other variables are zero. The presence of a Republican majority and the constant both had increased significance due to the inclusion of additional variables in the full model attenuating their explanatory power.

3. Random Forest Classification

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has participated in initiatives that are tracked by the UNFCCC. The proportion of the population categorised as white, the median household income, the lend temperature trend and the median age are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority, despite being significant in the forward selection model, the number of natural disasters and the land temperature trend were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasInitiativeParticipations	
	Importance	Rank
population	0.223	1
whiteProp	0.105	2
medianHouseholdIncome	0.088	3
tempDiff	0.084	4
medianAge	0.084	5
lessThanHighSchoolProp	0.083	6
unemploymentRate	0.083	7
avgEmissionsPerCapita	0.082	8
povertyProp	0.079	9
numDisasters	0.059	10
redCounty	0.030	11

a) Engagements – Climate Action Plan

The dependent variable 'hasClimateActionPlans' is positive for 74.1% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had two statistically significant coefficients, and none at the 1% level. A city's population and the median household income were significant at the 10% and 5% levels respectively and both have positive effects on the log odds of that city having climate action plans tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasClimateActionPlans	
	Full	Forward Selection
constant	-0.6963	2.9802
population	0.0000*	0.0000*
redCounty	-0.0495	
unemploymentRate	22.0077	
povertyProp	6.0801	
tempDiff	0.1765	
numDisasters	-0.0015	
avgEmissionsPerCapita	-0.0034	
lessThanHighSchoolProp	-8.9975	-8.4656**
medianHouseholdIncome	0.0000**	0.0000**
medianAge	-0.0949	-0.0839*
whiteProp	1.8492	
AIC	243	**231**
BIC	282	**248**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included four statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC climate action plan engagement compared to the other variables. The median age and the proportion of people with less than a high school diploma both had negative effects on log odds and had increased significance due to the inclusion of additional variables in the full model attenuating their explanatory power.

3. Random Forest Classification

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has climate action plans that are tracked by the UNFCCC. The proportion of the population categorised as white and the proportion of the population in poverty are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority, despite being significant in the forward selection model, the number of natural disasters and the land temperature trend were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasClimateActionPlans	
	Importance	Rank
population	0.293	1
medianHouseholdIncome	0.110	2
whiteProp	0.095	3
medianAge	0.083	4
povertyProp	0.078	5
lessThanHighSchoolProp	0.076	6
unemploymentRate	0.072	7
avgEmissionsPerCapita	0.060	8
numDisasters	0.059	9
tempDiff	0.058	10
redCounty	0.014	11

b) Climate Focus Actions – Mitigation

The dependent variable 'hasMitigations' is positive for 87.3% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had two statistically significant coefficients, and none at the 1% level. The median household income and the median age were significant at the 10% and 5% levels respectively and had opposite effects on the log odds of that city having mitigations tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasMitigations	
	Full	Forward Selection
constant	1.5733	2.4879
population	0.0000	0.0000
redCounty	0.6430	
unemploymentRate	29.5499	
povertyProp	-2.4839	
tempDiff	0.3517	
numDisasters	-0.0033	-0.0070***
avgEmissionsPerCapita	0.0442	
lessThanHighSchoolProp	-7.7900	
medianHouseholdIncome	0.0001*	0.0000***
medianAge	-0.1556**	-0.1128**
whiteProp	1.4624	
AIC	187	**179**
BIC	226	**195**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included three statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC mitigations compared to the other variables, however, it was insignificant. The median household income and the number of natural disasters both had increased significance due to the inclusion of additional variables in the full model attenuating their explanatory power.

3. Random Forest Classification

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has adaptations that are tracked by the UNFCCC. The proportion of the population categorised as white and the proportion of the population in poverty are ranked in the top five most important variables despite being insignificant in the logistic regression models. The presence of a Republican majority, the average emissions per capita and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasMitigations	
	Importance	Rank
population	0.233	1
medianAge	0.148	2
medianHouseholdIncome	0.116	3
whiteProp	0.102	4
povertyProp	0.083	5
lessThanHighSchoolProp	0.081	6
unemploymentRate	0.061	7
tempDiff	0.059	8
avgEmissionsPerCapita	0.059	9
numDisasters	0.051	10
redCounty	0.008	11

b) Climate Focus Actions – Adaptation
The dependent variable 'hasAdaptations' is positive for 78.2% of observations used in the models.

1. Logistic Regression
The logistic regression model of all observations had one statistically significant coefficients, and none at the 1% level. The presence of a Republican majority was significant at the 10% level and had a negative effect on the log odds of that city had adaptations tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasAdaptations	
	Full	Forward Selection
constant	0.7971	-3.1342***
population	0.0000	0.0000
redCounty	-0.7186*	
unemploymentRate	1.2851	
povertyProp	-7.5449	
tempDiff	0.1111	
numDisasters	-0.0034	
avgEmissionsPerCapita	0.0450	0.0573
lessThanHighSchoolProp	1.3696	
medianHouseholdIncome	0.0000	0.0000***
medianAge	-0.0748	
whiteProp	1.0639	
AIC	229	**220**
BIC	268	**233**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection
The forward selection model included two statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. Population was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC adaptations compared to the other variables, however, it was insignificant. The median household income and the constant both had increased significance to the 1% level due to the inclusion of additional variables in the full model attenuating their explanatory power. The average emissions per capita was included but was not significant. The presence of a Republican majority was not selected despite being significant in the full model.

3. Random Forest Classification
The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has mitigations that are tracked by the UNFCCC. All of the features ranked in the top five most important variables were not insignificant in the logistic regression models. The presence of a Republican majority, the land temperature trend and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasAdaptations	
	Importance	Rank
population	0.280	1
medianHouseholdIncome	0.129	2
medianAge	0.098	3
povertyProp	0.091	4
whiteProp	0.091	5
lessThanHighSchoolProp	0.071	6
unemploymentRate	0.063	7
avgEmissionsPerCapita	0.059	8
tempDiff	0.057	9
numDisasters	0.048	10
redCounty	0.014	11

b) Climate Focus Actions – Finance
The dependent variable 'hasFinanceActions' is positive for 9.6% of observations used in the models.

1. Logistic Regression
The logistic regression model of all observations had two statistically significant coefficients, and none at the 1% level. The number of natural disasters and the proportion of people with less than a high school diploma were significant at the 5% and 10% levels respectively and had opposite effects on the log odds of that city having finance actions tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasFinanceActions	
	Full	Forward Selection
constant	5.3011	-2.7623***
population	0.0000	0.0000
redCounty	0.4428	
unemploymentRate	78.4613	
povertyProp	-7.0380	
tempDiff	0.3943	
numDisasters	0.0069**	
avgEmissionsPerCapita	-0.0603	
lessThanHighSchoolProp	-28.9080*	
medianHouseholdIncome	0.0000	
medianAge	-0.0901	
whiteProp	-2.7961	
AIC	115	**104**
BIC	154	**111**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection
The forward selection model included one statistically significant coefficient. This model offers a better fit with a lower AIC and BIC. Population was the first, and only, additional variable selected which possibly indicates its relative importance in explaining UNFCCC adaptations compared to the other variables, however, it was insignificant. The constant had increased significance to the 1% level due to the inclusion of additional variables in the full model attenuating their explanatory power. The number of natural disasters and the proportion of people with less than a high school diploma were not selected despite being significant in the full model.

3. Random Forest Classification
The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has mitigations that are tracked by the UNFCCC. All of the features ranked in the top five most important variables were not insignificant in the logistic regression models, except for the proportion of people with less than a high school diploma. The presence of a Republican majority, the average emissions per capita and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasFinanceActions	
	Importance	Rank
population	0.246	1
whiteProp	0.109	2
unemploymentRate	0.100	3
medianAge	0.097	4
lessThanHighSchoolProp	0.096	5
tempDiff	0.087	6
medianHouseholdIncome	0.076	7
povertyProp	0.072	8
numDisasters	0.065	9
avgEmissionsPerCapita	0.039	10
redCounty	0.015	11

c) Progress Tracking – Emissions Inventory

The dependent variable 'hasEmissionInventory' is positive for 62.9% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had five statistically significant coefficients, and three at the 1% level. A city's population, the proportion of its population which is in poverty and the median household income all were significant at 1% level and have positive effects on the log odds of that city having an emissions inventory that is tracked by the UNFCCC. The proportion of a city's population with less than a high school diploma was significant at the 10% level and resulted in decreases in the log odds of having a UNFCCC emissions inventory. The constant was significant at the 5% level indicating a baseline log odds of -8.3588 when all other variables are zero. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasEmissionInventory	
	Full	Forward Selection
constant	-8.3588**	-6.5556***
population	0.0000***	0.0000***
redCounty	-0.3825	
unemploymentRate	20.3213	
povertyProp	22.8593***	20.8728***
tempDiff	0.2441	
numDisasters	-0.0006	
avgEmissionsPerCapita	0.0134	
lessThanHighSchoolProp	-12.0784**	-11.1956**
medianHouseholdIncome	0.0001***	0.0001***
medianAge	-0.0315	
whiteProp	1.6235	
AIC	242	**231**
BIC	281	**247**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included five statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. All the variables were significant in the full model at the same level except for the constant which increased in significance due to the inclusion of additional variables in the full model attenuating their explanatory power. Population was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC emissions inventory tracking compared to the other variables.

3. Random Forest Classification

The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has emissions inventories that are tracked by the UNFCCC. The median age and the proportion of a city's population which is categorised as white are both ranked in the top five most important variables but were not insignificant in the logistic regression models. The presence of a Republican majority, the land temperature trend and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasEmissionInventory	
	Importance	Rank
population	0.278	1
lessThanHighSchoolProp	0.108	2
medianAge	0.103	3
whiteProp	0.102	4
medianHouseholdIncome	0.100	5
povertyProp	0.077	6
unemploymentRate	0.069	7
avgEmissionsPerCapita	0.053	8
numDisasters	0.050	9
tempDiff	0.048	10
redCounty	0.012	11

c) Progress Tracking – Risk Assessment
The dependent variable 'hasRiskAssessments' is positive for 61.4% of observations used in the models.

1. Logistic Regression
The logistic regression model of all observations had two statistically significant coefficients, both at the 1% level. A city's population and the median household income were both significant at the 1% level and had positive effects on the log odds of that city having risk assessments that are tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasRiskAssessments	
	Full	Forward Selection
constant	-3.8236	-2.7077***
population	0.0000***	0.0000***
redCounty	-0.2678	
unemploymentRate	28.2781	
povertyProp	6.9209	
tempDiff	0.3567	
numDisasters	-0.0018	
avgEmissionsPerCapita	0.0004	
lessThanHighSchoolProp	-4.7900	
medianHouseholdIncome	0.0001***	0.0000***
medianAge	-0.0718	
whiteProp	1.2963	
AIC	256	**245**
BIC	296	**255**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection
The forward selection model included three statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. All the variables were significant in the full model at the same level except for the constant which increased in significance due to the inclusion of additional variables in the full model attenuating their explanatory power. Population was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC risk assessment tracking compared to the other variables.

3. Random Forest Classification
The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has risk assessments that are tracked by the UNFCCC. The median household income, which was the only other significant variable in the full model is ranked second in feature importance. The median age, the proportion of people with less than a high school diploma and the proportion of a city's population which is categorised as white are ranked in the top five most important variables but were not insignificant in the logistic regression models. The presence of a Republican majority, the average emissions per capita and the number of natural disasters were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasRiskAssessments	
	Importance	Rank
population	0.247	1
medianHouseholdIncome	0.124	2
whiteProp	0.100	3
medianAge	0.087	4
lessThanHighSchoolProp	0.081	5
tempDiff	0.080	6
povertyProp	0.078	7
unemploymentRate	0.074	8
numDisasters	0.065	9
avgEmissionsPerCapita	0.055	10
redCounty	0.009	11

c) Progress Tracking – Impact

The dependent variable 'hasImpact' is positive for 6.1% of observations used in the models.

1. Logistic Regression

The logistic regression model of all observations had two statistically significant coefficients, neither at the 1% level. The number of natural disasters and the proportion of people with less than a high school diploma were significant at 5% and 10% levels respectively and had opposite effects on the log odds of that city having impact that is tracked by the UNFCCC. All other variables had statistically insignificant impacts on the dependent variable.

Dependent Variable	hasImpact	
	Full	Forward Selection
constant	-9.0717	-4.7994**
population	0.0000	
redCounty	0.1913	
unemploymentRate	9.5872	
povertyProp	24.4481	19.4192
tempDiff	-0.0505	
numDisasters	0.0121**	0.0132**
avgEmissionsPerCapita	0.0069	
lessThanHighSchoolProp	-20.5532*	-25.3963*
medianHouseholdIncome	0.0000	
medianAge	-0.0529	
whiteProp	3.8310	
AIC	92	**78**
BIC	131	**91**

** p<0.05, ** p<0.01, *** p<0.001*

2. Forward Selection

The forward selection model included three statistically significant coefficients. This model offers a better fit with a lower AIC and BIC. All the variables were significant in the full model at the same level except for the constant which increased in significance due to the inclusion of additional variables in the full model attenuating their explanatory power. The number of natural disasters was the first additional variable selected which possibly indicates its relative importance in explaining UNFCCC impact tracking compared to the other variables. The proportion of a city's population which is in poverty is included in the model despite being insignificant.

3. Random Forest Classification

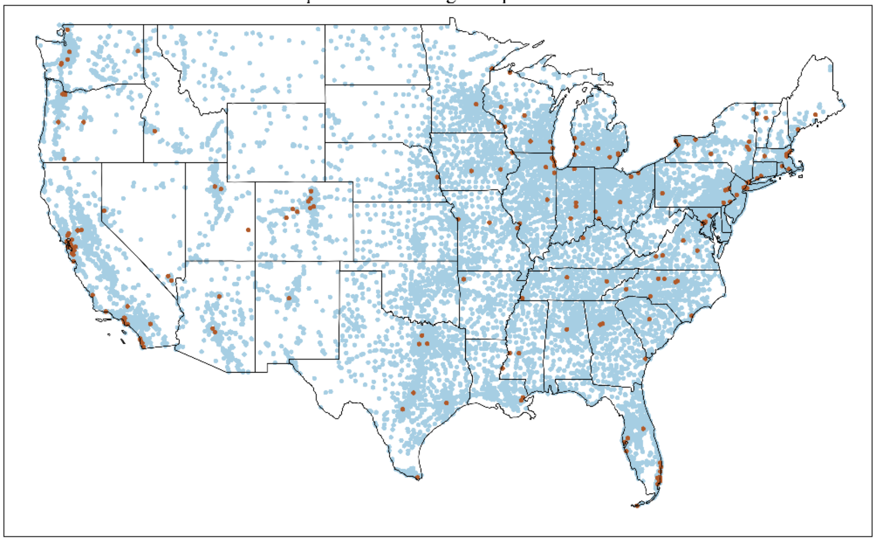
The random forest feature importances ranked a city's population as the variable which contributed the most to nodal purity in the classification of whether the city has risk assessments that are tracked by the UNFCCC. All the variables ranked in the top five most important variables were not insignificant in the logistic regression models. The presence of a Republican majority, the average emissions per capita and the number of natural disasters, despite being significant in the previous models, were least important and therefore, accounted for the least increase in the Gini index.

Feature	hasImpact	
	Importance	**Rank**
population	0.267	1
povertyProp	0.132	2
medianAge	0.103	3
medianHouseholdIncome	0.095	4
unemploymentRate	0.092	5
lessThanHighSchoolProp	0.087	6
whiteProp	0.077	7
tempDiff	0.049	8
numDisasters	0.048	9
avgEmissionsPerCapita	0.046	10
redCounty	0.006	11

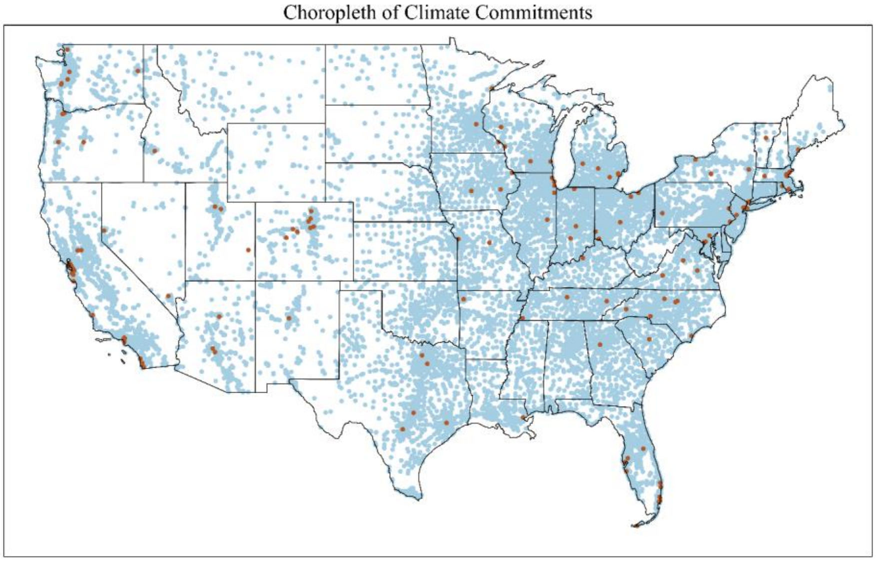
Exploratory Data Analysis

Choropleths

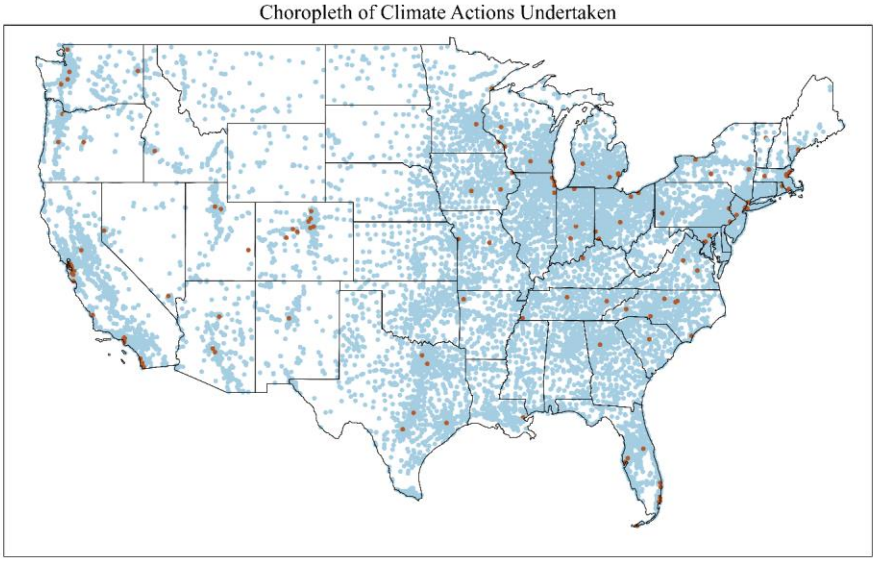
Choropleth of Cities Signed up to UNFCCC



Choropleth of Climate Commitments

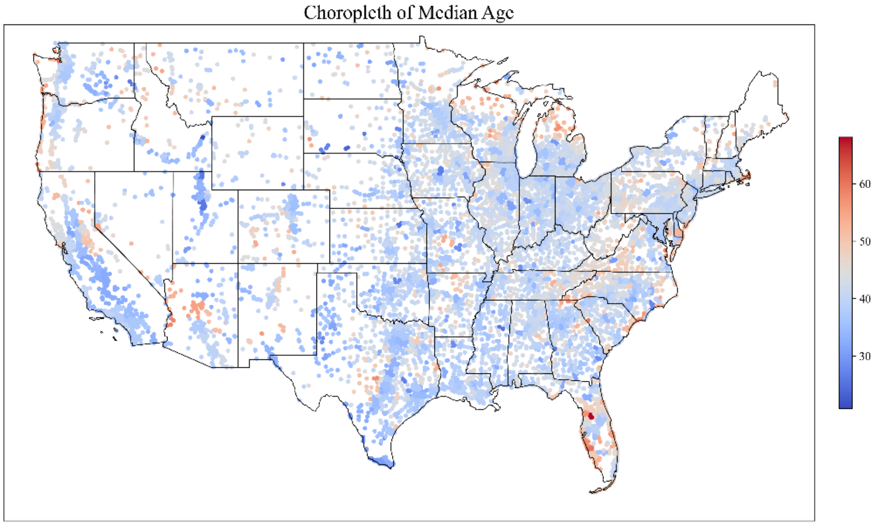


Choropleth of Climate Actions Undertaken

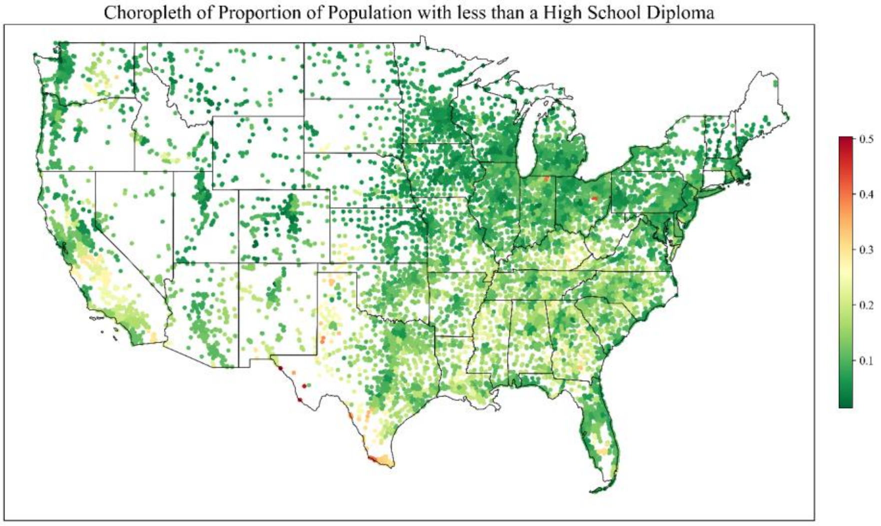


Demographic Features

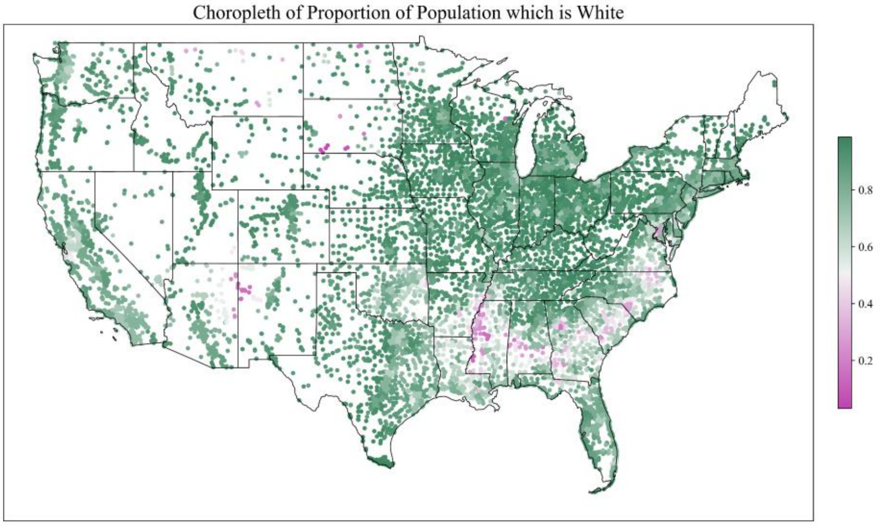
Choropleth of Median Age



Choropleth of Proportion of Population with less than a High School Diploma

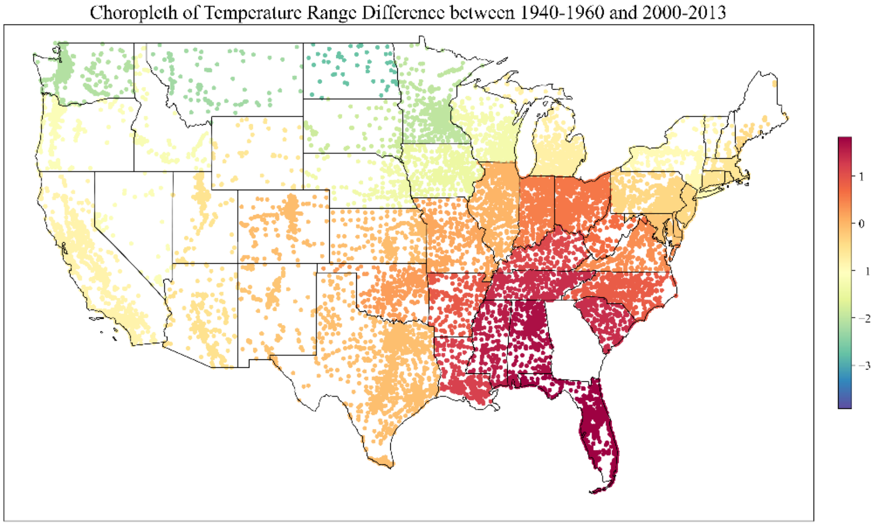


Choropleth of Proportion of Population which is White

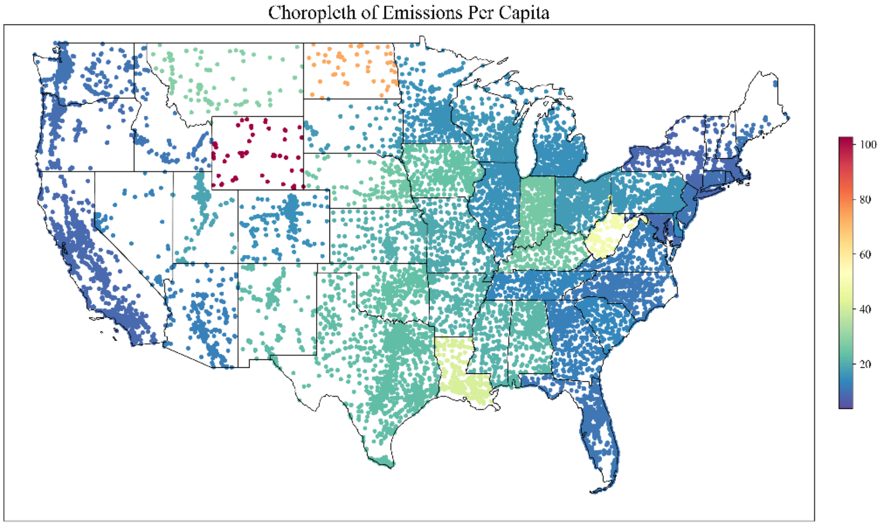


Climate Change Features

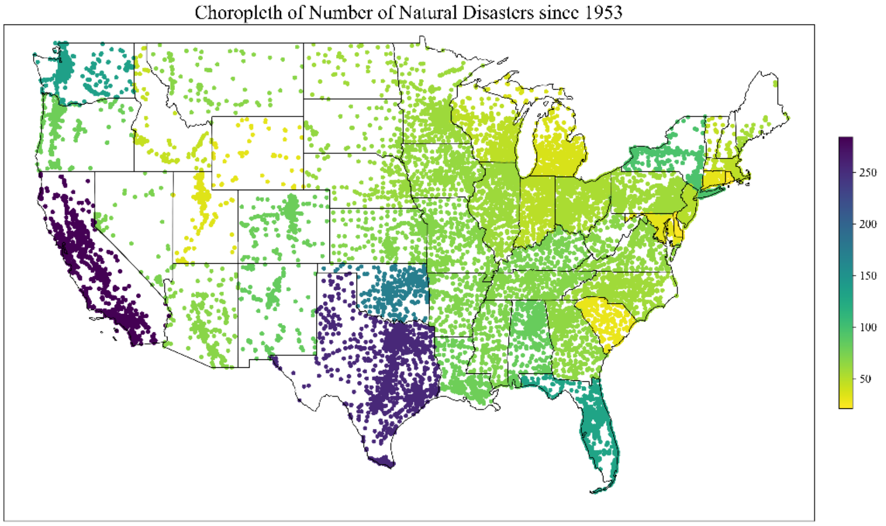
Choropleth of Temperature Range Difference between 1940-1960 and 2000-2013



Choropleth of Emissions Per Capita

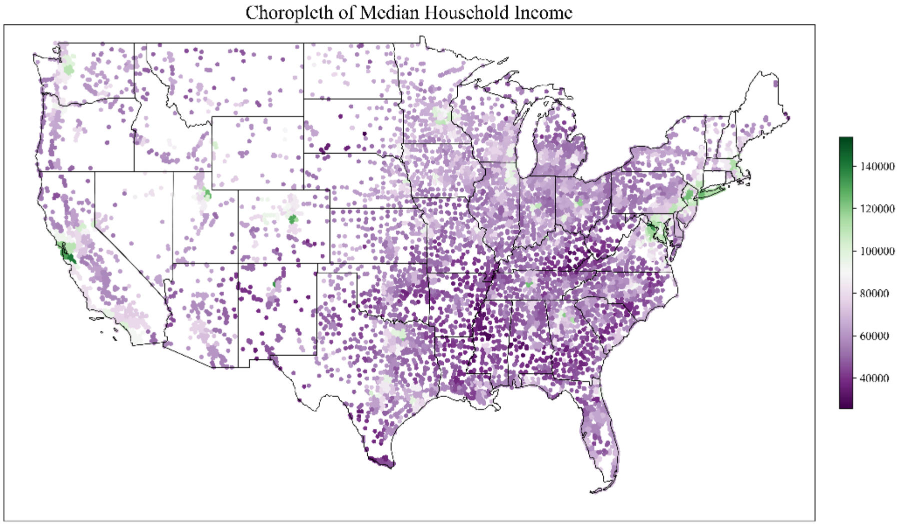


Choropleth of Number of Natural Disasters since 1953

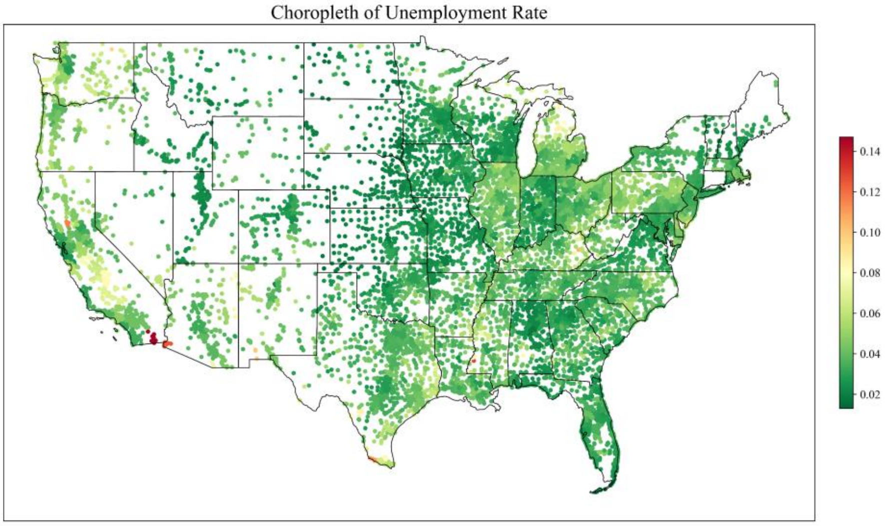


Economic Features

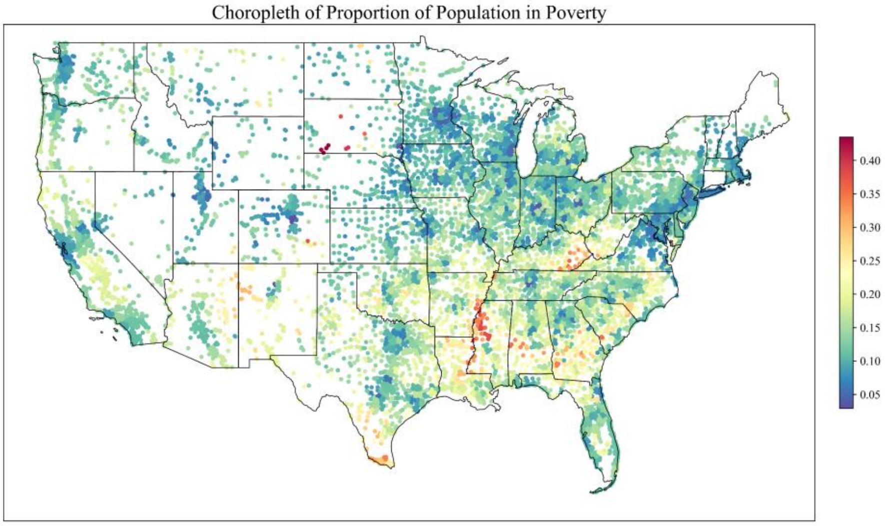
Choropleth of Median Household Income



Choropleth of Unemployment Rate

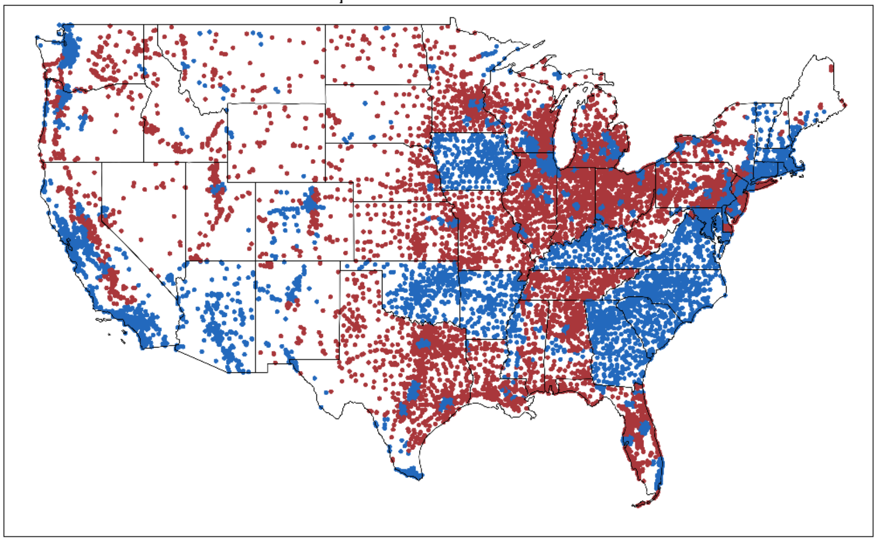


Choropleth of Proportion of Population in Poverty

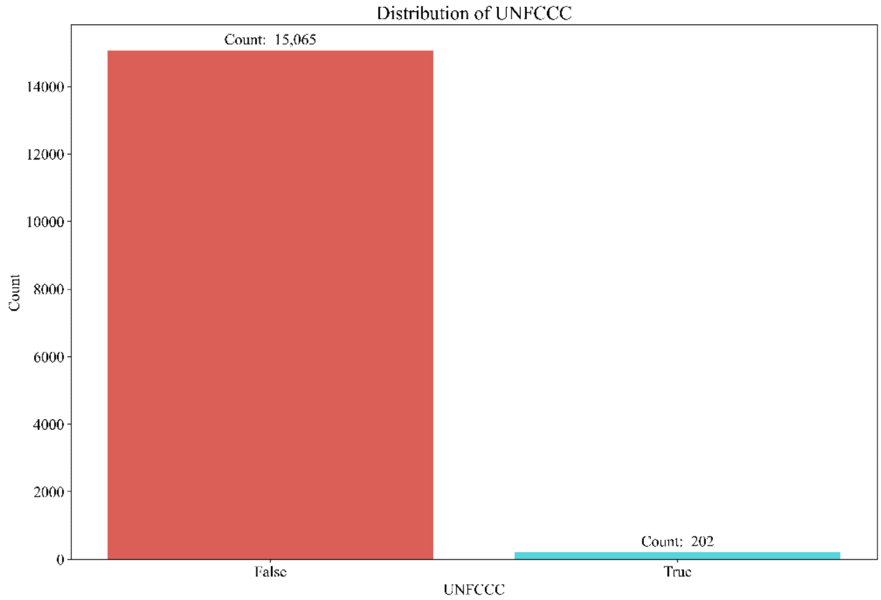


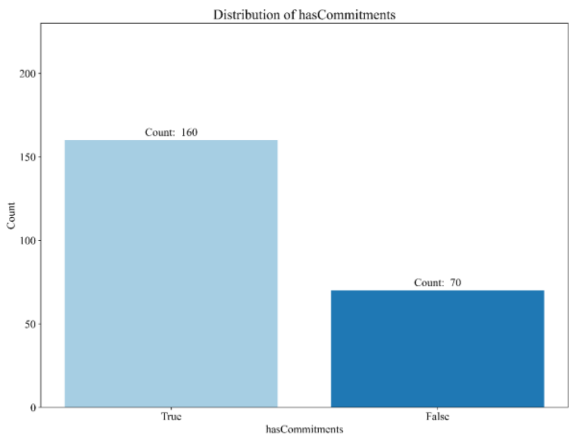
Political Feature

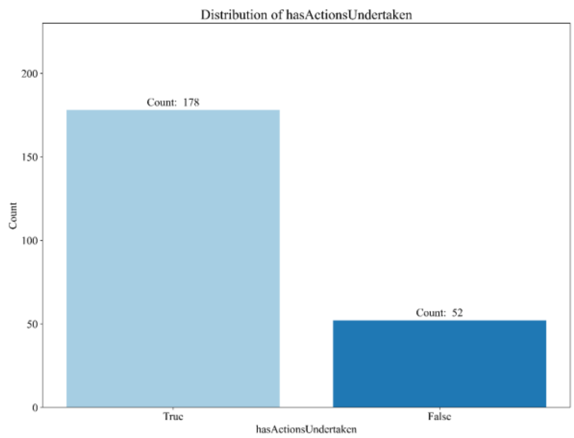
Choropleth of 2020 Election Results

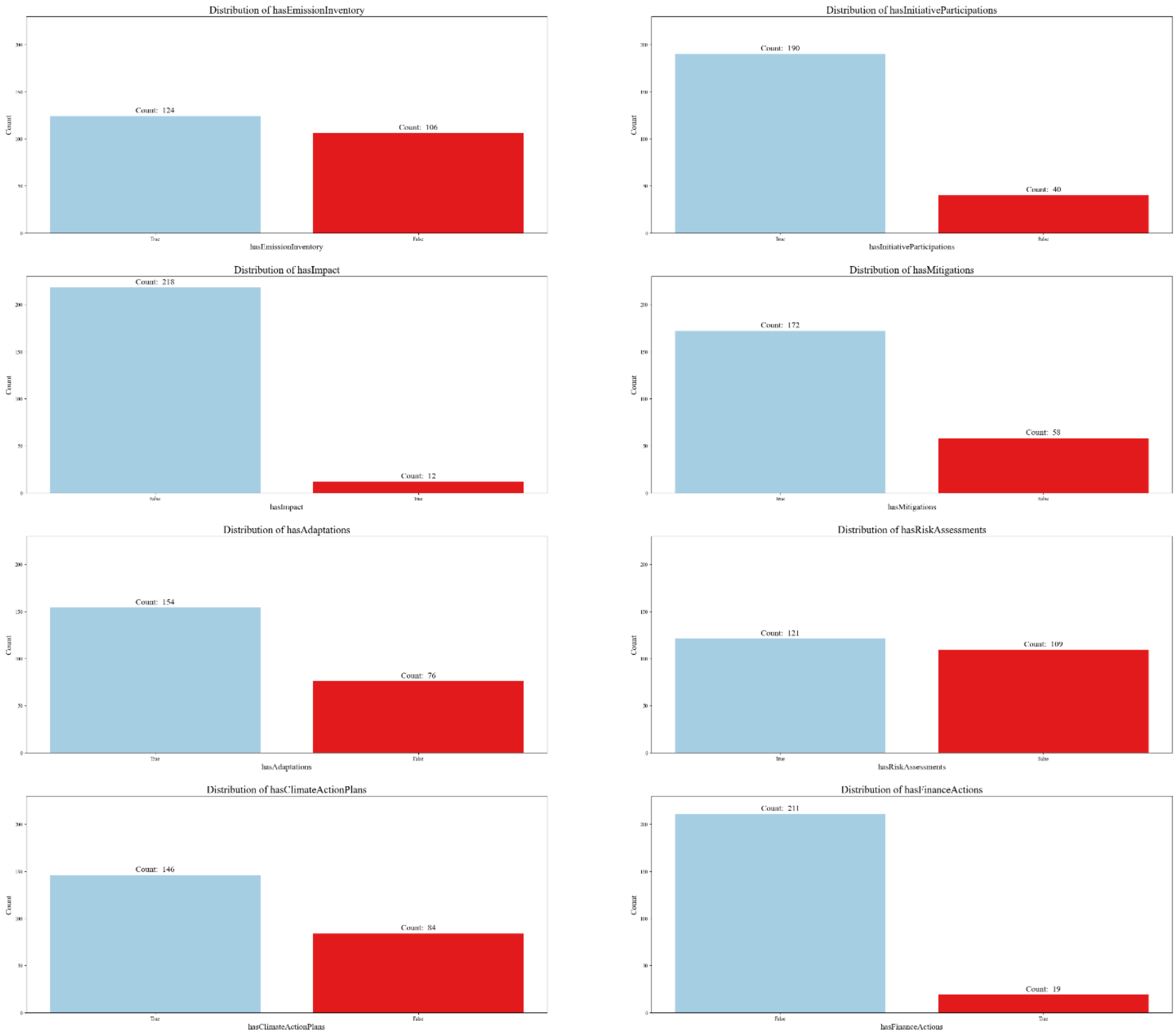


Distribution Plots



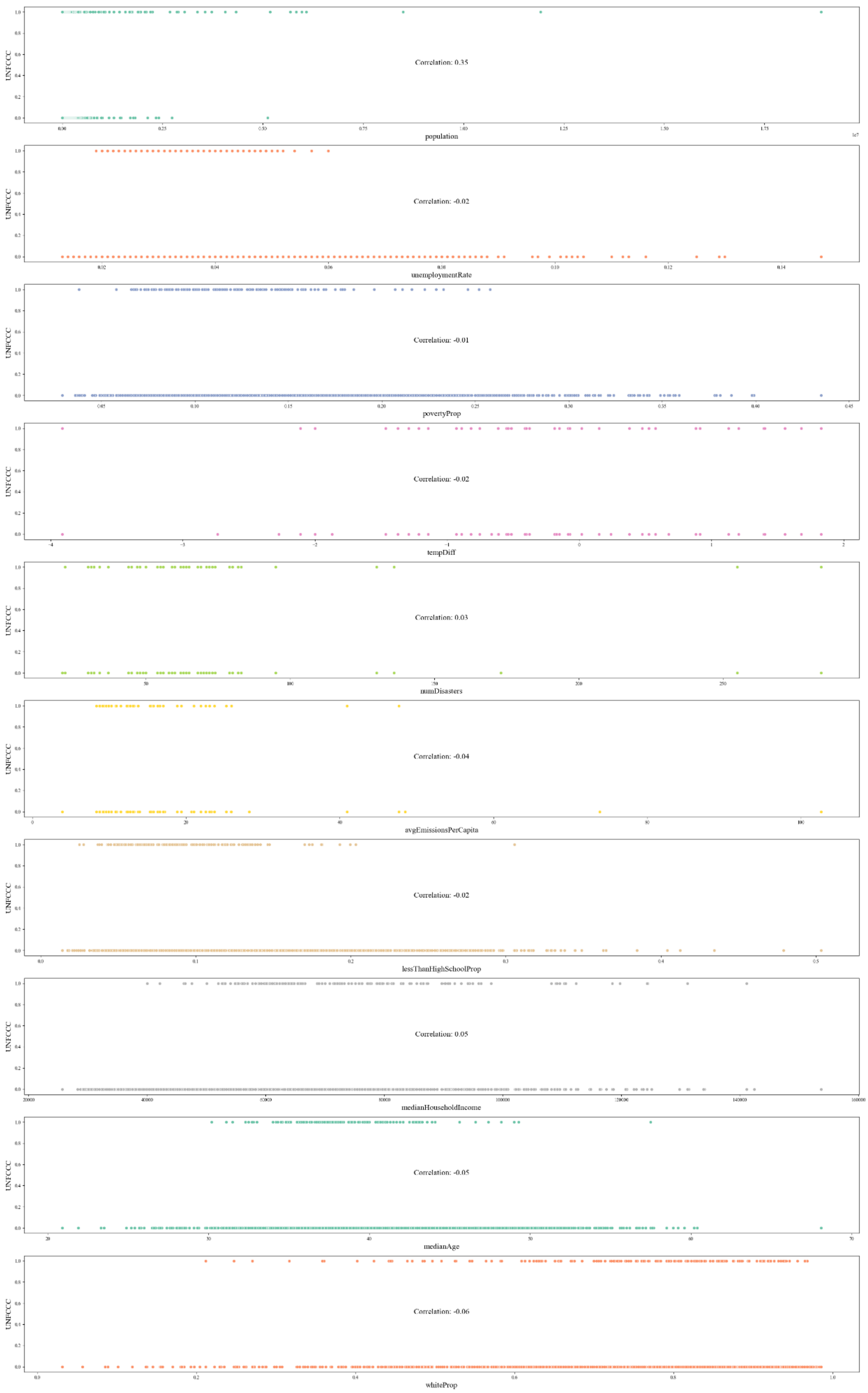




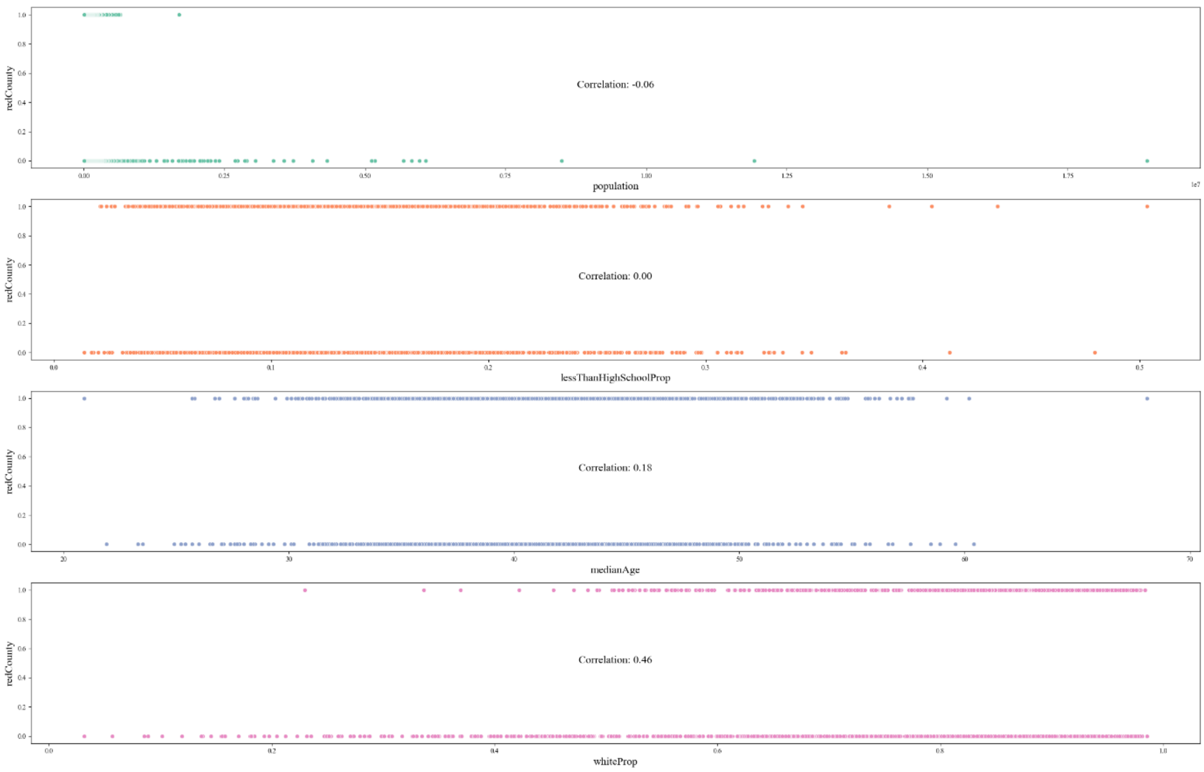

Distribution of hasEmissionInventory
Count: 124
Count: 106
Count
hasEmissionInventory
Distribution of hasInitiativeParticipations
Count: 190
Count: 40
Count
hasInitiativeParticipations
Distribution of hasImpact
Count: 218
Count: 12
Count
hasImpact
Distribution of hasMitigations
Count: 172
Count: 58
Count
hasMitigations
Distribution of hasAdaptations
Count: 154
Count: 76
Count
hasAdaptations
Distribution of hasRiskAssessments
Count: 121
Count: 109
Count
hasRiskAssessments
Distribution of hasClimateActionPlans
Count: 146
Count: 84
Count
hasClimateActionPlans
Distribution of hasFinanceActions
Count: 211
Count: 19
Count
hasFinanceActions

Correlation Plots

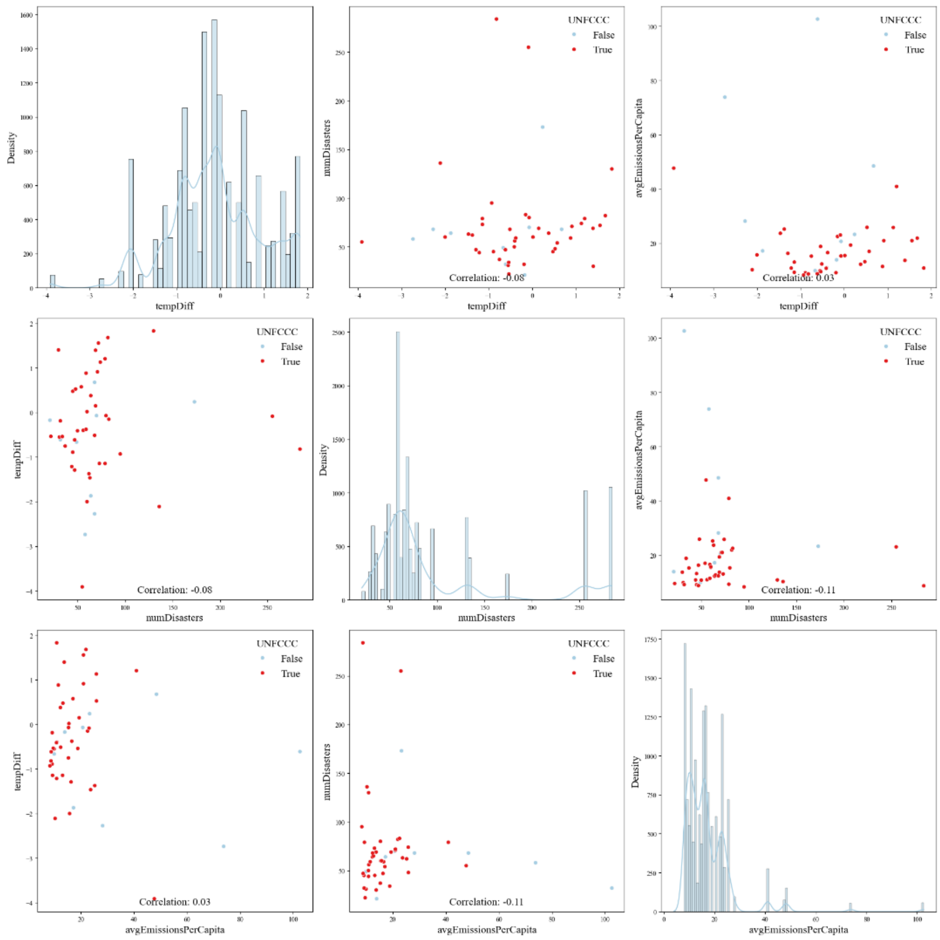
Correlation of variables of interest with the research question 1 target variable: 'UNFCCC'



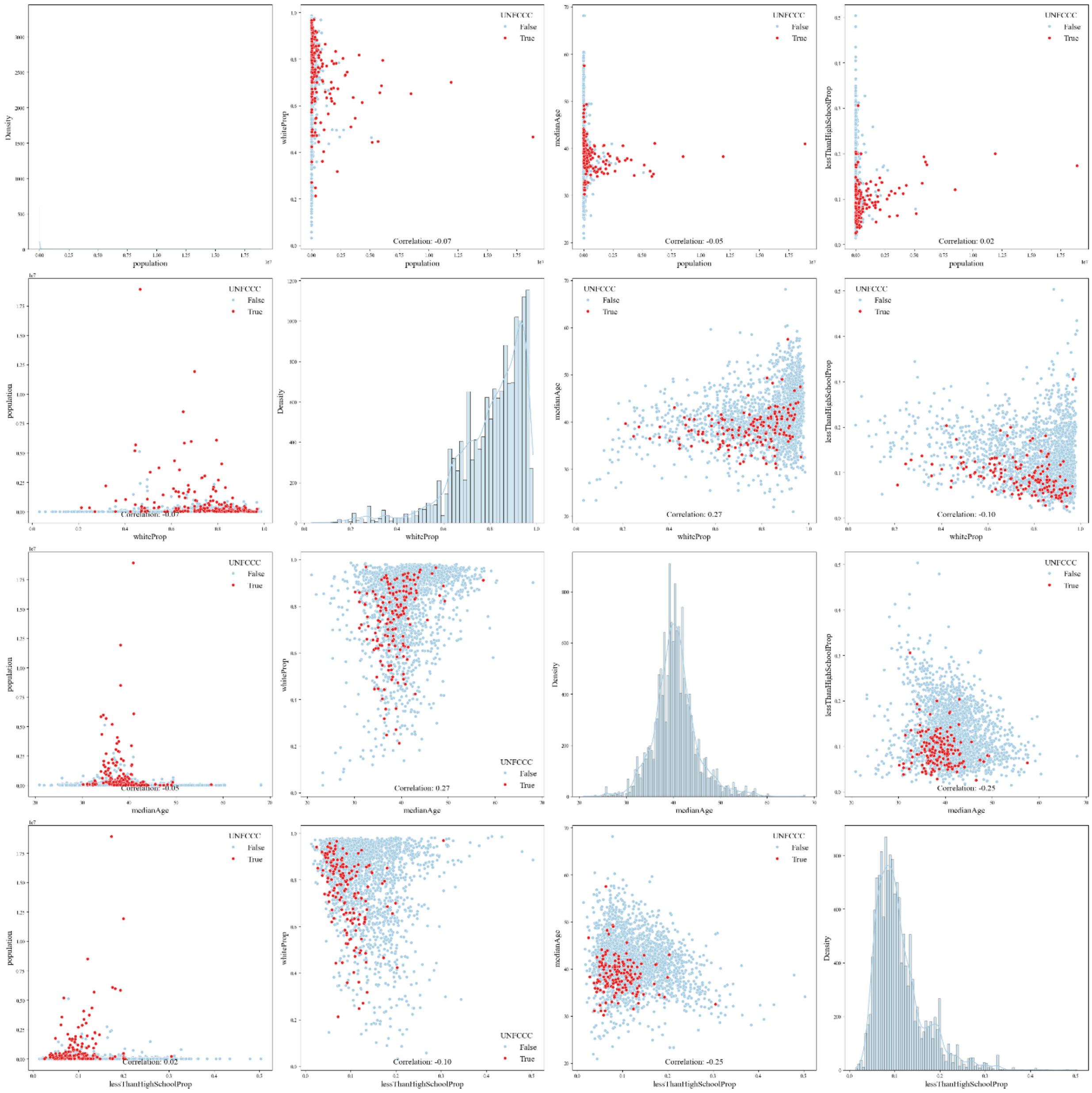
Correlation of Demographic Features with the Political Feature



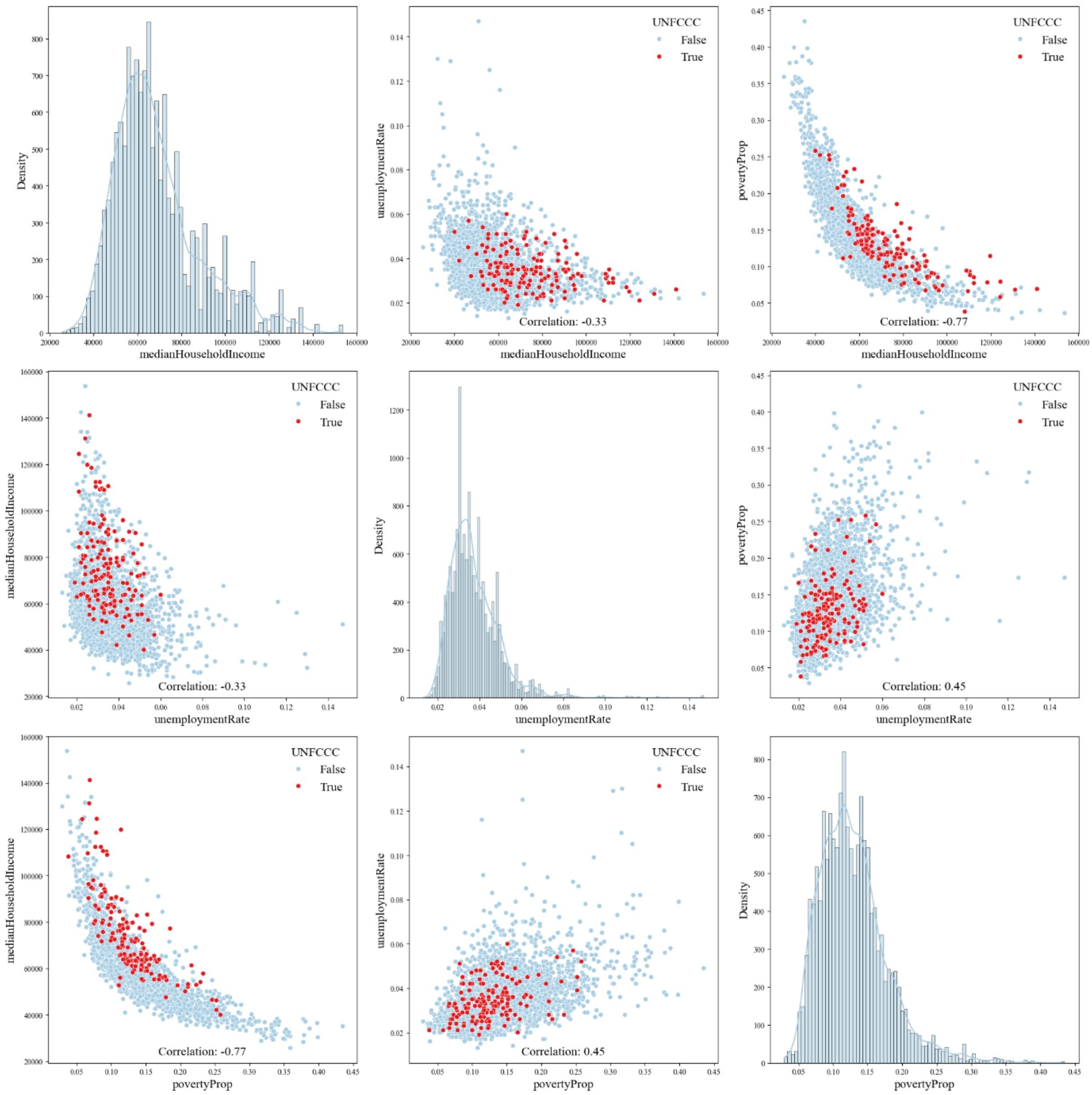
Correlation of Climate Change Features



Correlation of Demographic Features



Correlation of Economic Features



Methodology

Detailed descriptions of all statistical techniques used in the analysis can be found below.

1. Logistic Regression

   Logistic regression is a statistical model which uses maximum likelihood estimators to obtain the coefficients of the following logit model:

$$P(X) = \frac{e^{\beta_0 + \sum_{i=1}^{p} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{p} \beta_i X_i}}$$

   where $P(X)$ is the probability of an observation being equal to 1.

   The advantage of the logistic model over conventional linear regression is due to the logit function's sigmoid shape ensuring probability estimates that lie on the interval from 0 to 1. The function is bounded by these limits and therefore, is appropriate for modelling binary outcomes like the target variables in the data set of interest.

   The interpretation of the logistic regression coefficients is different to that of linear regression. A coefficient of β_i, for $i \geq 1$ means that for each one unit increase in X_i, the log odds change by β_i. This can be seen by rearranging the logit model equation to obtain

$$\ln\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i$$

   In this analysis, the statistical significance of the variables' coefficients is used to determine the usefulness of those variables in explaining the variation in the target variables.

2. Forward Selection

   Forward selection is a stepwise modelling technique that is used to select independent variables. The algorithm works by, in the case of this analysis, starting with a logistic regression of the dependent variable and a constant as a benchmark. Then it iterates through the other potential independent variables each time adding it to the benchmark model. If the model with the lowest Akaike Information Criterion (AIC) is lower than the benchmark model, indicating a better fit, it becomes the new benchmark. This process is continued, adding a new variable each time until the model with the minimum AIC is not lower than the benchmark model.

   Akaike Information Criterion (AIC) is calculated using the following formula

$$AIC = 2k - 2\ln(L)$$

   where k is the number of estimated parameters in the model and L is the likelihood value of the mode calculated during the maximum likelihood estimation of the model.

   Some insights surrounding feature importance can be derived from the final model, however, the order variables were chosen as the best variable given the previous features in the model. Therefore, variables selected later in the process may still be more important.

3. Random Forest Classification

   Random forests are ensemble models that combine the classification of a high number of basic classification trees to reduce the variance of the singular trees. A basic classification tree partitions the feature space linearly, choosing splits that improve Gini impurity. The Gini index is as follows:

$$G = \sum_{k=1}^{K} \hat{p}_k(1 - \hat{p}_k)$$

   Random forested fit a high number of decision trees but to avoid each tree being the same, at each split the trees are limited to only using $\sqrt{V}$ random variables, where V is the total number of variables. The overall classification of each observation is based on a majority rules approach across all the trees in the random forest.

To assess the deterministic power of the variables, feature importance is used. This is calculated by averaging the total amount that the Gini index is decreased in splits using a given variable over all the trees in the random forest. The mean decrease is expressed relative to the maximum decrease to scale all the feature importances. This measure of feature importance does not give any indication of the direction of the effect, as can inferred from logistic regression coefficients.

Data

All data, sources and their corresponding variables, used in the analysis can be found detailed in the table below.

Dataset	Description	Variables of Interest	Sources
UNFCCC	Global Climate Action actor tracking data on all cities in the US that are signed up to the UNFCCC	hasCommitments, hasActionsUndertaken, hasEmissionInventory, hasInitiativeParticipations, hasImpact, hasMitigations, hasAdaptations, hasRiskAssessments, hasClimateActionPlans, hasFinanceActions	Global Climate Action UNFCCC - Actor Tracking (2022).
All US Cities	Geographical data on all cities in the US.	city, state, state_name, fips, county, latitude, longitude, population	US Cities Database | Simplemaps.com.
2020 Election Data	2020 US Presidential Election data at both State and County level	redCounty, redState	MIT Election Data and Science Lab, 2018, "County Presidential Election Returns 2000-2020", Wikipedia Contributors (2019). 2020 United States presidential election.
Land Temperature Data	State level monthly average temperatures from 1829 to 2013.	tempDiff	www.kaggle.com. (Berkeley Earth). Climate Change: Earth Surface Temperature Data.
Natural Disaster Data	Total number of natural disasters for each State since 1953	numDisasters	worldpopulationreview.com. Natural Disasters by State
Emissions Data	State level energy-related emissions per capita, 2016-2021	avgEmissionsPerCapita	U.S. Energy Information Administration, State Energy Data System and EIA calculations.
Education Level Data	County level data on proportions of populations to have obtained specific levels of education	lessThanHighSchoolProp	USDA, Economic Research Service using data from U.S. Department of Commerce, Bureau of the Census, 1970, 1980, 1990, 2000 Censuses of Population, and 2008–12 and 2017–21 American Community Survey 5-year period county-level estimates.
Household Income	County level data on the median household income, 2021	medianHouseholdIncome	U.S. Department of Commerce, Bureau of the Census, Small Area Income and Poverty Estimates (SAIPE) Program.
Unemployment Rate	County-Level Unemployment Rate Data, 2022	unemploymentRate	Unemployment: U.S. Department of Labor, Bureau of Labor Statistics, Local Area Unemployment Statistics (LAUS)
Demographic Features	County-Level Demographic Statistics	whiteProp, medianAge	U.S. Census Bureau, Population Division - Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2022
Poverty Level Data	County level data estimating the level of poverty, 2021	povertyProp	U.S. Department of Commerce, Bureau of the Census, Small Area Income and Poverty Estimates (SAIPE) Program.

Details of the calculations used to create the variables of interest that were transformed from their original sources can be found in the table below.

Variables of Interest	Calculation
redCounty	At a county-level, redCounty was assigned to be True if the total number of votes to the Republican candidate (Trump / Pence) in the 2020 Election was greater than that of the Democrat candidate (Biden / Harris).
tempDiff	For each state in each year the differences between July and January average land temperatures were calculated. Then for each state, the average of the following two periods was calculated 1. 1940 – 1960 2. 2000 – 2013 and the tempDiff variable for each state was the difference between these two averages (2000 to 2013 minus 1940 to 1960). The objective was to get an indication of the average temperature trend in each state. Whether it was rising, falling or not changing potentially due to climate change. 2013 was the latest available year of data.
avgEmissionsPerCapita	Average energy-related emissions of each state from 2016 to 2021. An average was used to mitigate the effects of pandemic-related anomalous values.

Bibliography

Chandrasekharan, S., Sreedharan, J. and Gopakumar, A., 2019. Statistical issues in small and large sample: need of optimum upper bound for the sample size. International Journal of Computational and Theoretical Statistics, 6(2).

Fagan, M. and Huang, C. (2019). A look at how people around the world view climate change. [online] Pew Research Center. Available at: https://www.pewresearch.org/short-reads/2019/04/18/a-look-at-how-people-around-the-world-view-climate-change/.

Global Climate Action UNFCCC - Actor Tracking (2022). Available at: https://climateaction.unfccc.int/Actors.

Hickman, C., Marks, E., Pihkala, P., Clayton, S., Lewandowski, R.E., Mayall, E.E., Wray, B., Mellor, C. and Van Susteren, L. (2021). Climate anxiety in children and young people and their beliefs about government responses to climate change: a global survey. The Lancet Planetary Health, 5(12), pp.e863-e873.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

MIT Election Data and Science Lab (2018), "County Presidential Election Returns 2000-2020", https://doi.org/10.7910/DVN/VOQCHQ, Harvard Dataverse, V12

National Association of Counties (NACo) (2012). History of County Government Part 1. Available at: https://www.nvnaco.org/wp-content/uploads/History-and-Overview-of-County-Government-n-the-U.S.-NACo.pdf

R. Wilms, E. Mäthner, L. Winnen, R. Lanwehr, (2021), Omitted variable bias: A threat to estimating causal relationships, Methods in Psychology, Volume 5, 100075, ISSN 2590-2601, https://doi.org/10.1016/j.metip.2021.100075.

simplemaps.com. US Cities Database | Simplemaps.com. [online] Available at: https://simplemaps.com/data/us-cities.

Tyson, A., Funk, C. and Kennedy, B. (2023). What the data says about Americans' views of climate change. [online] Pew Research Center. Available at: https://www.pewresearch.org/short-reads/2023/08/09/what-the-data-says-ab

U.S. Department of Commerce, Bureau of the Census, Small Area Income and Poverty Estimates (SAIPE) Program.

U.S. Department of Labor, Bureau of Labor Statistics, Local Area Unemployment Statistics (LAUS).

U.S. Energy Information Administration, State Energy Data System and EIA calculations. Available at: https://www.eia.gov/environment/emissions/state/.

United Nations (1992). United Nations Framework Convention on Climate Change. [online] Available at: https://unfccc.int/files/essential_background/background_publications_htmlpdf/application/pdf/conveng.pdf

USDA, Economic Research Service using data from U.S. Department of Commerce, Bureau of the Census, 1970, 1980, 1990, 2000 Censuses of Population, and 2008–12 and 2017–21 American Community Survey 5-year period county-level

Wikipedia Contributors (2019). 2020 United States presidential election. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/2020_United_States_presidential_election.

worldpopulationreview.com. Natural Disasters by State [Updated May 2023]. [online] Available at: https://worldpopulationreview.com/state-rankings/natural-disasters-by-state.

www.kaggle.com. (Berkeley Earth). Climate Change: Earth Surface Temperature Data. [online] Available at: https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data.