US State Temperature Analysis

Nimrat Grewal

ECO 225

April 16, 2023

Introduction

In this paper, the economic question I would like to answer is the following: How have the temperatures across states in the US changed overtime and what patterns can be observed in different regions? Is there a correlation between temperature patterns and economic indicators in different regions? What factors may contribute to these changes?

There are a few papers that have talked about this topic extensively and gone into further details concerning the economic variables they have studied. For example, a paper called "Temperature and growth: A panel analysis of the United States." [1] in 2019, assessed state level temperature data for the states using 1957-2012 historic data to check its effect on state level output, and suggested that rising temperatures may cause state level economic activity to decrease, significantly during the summer. They talked expensively about industry output, labour productivity and even conducted a regional analysis for temperature.

Next, another paper called "The identification of 10-to 20-year temperature and precipitation fluctuations in the contiguous United States."[2] studied climate fluctuations over 10-20 years to determine which regions produced the sharpest transitions onto the next, conducted under different simulations such as controlling for seasonality fluctuations in temperature and precipitation. They discovered that the most significant temperature fluctuations of 10-20 years have been associated with a temperature increase of 2 degrees Celsius or more during the winter and summer seasons.

1. Colacito, Riccardo, Bridget Hoffmann, and Toan Phan. "Temperature and growth: A panel analysis of the United States." Journal of Money, Credit and Banking 51.2-3 (2019): 313-368.

2. Karl, Thomas R., and William E. Riebsame. "The identification of 10-to 20-year temperature and precipitation fluctuations in the contiguous United States." Journal of Applied Meteorology and Climatology 23.6 (1984): 950-966.

Both of these papers have important implications based on their findings. Since the second paper suggests that climate fluctuations are more severe during the winter and summer months, combining with the results of the first paper, policymakers can properly identify that more severe increases in temperature over the summer overtime can cause aggregate state level output to drop slowly overtime. In this way, economic decisions can be impacted by the results of temperature studies.

My paper adds a unique implementation of temperature analysis for several reasons. First my temperature analysis for studying the temperature change overtime differs from these two papers. While the second paper uses 10-20 year fluctuations to study any trends, I have used a historic 50 year baseline average per state using my data and compared that against the rest of the temperature figured until 2013 to see how to temperature has changed compared the the historic average. The baseline average makes for a more meaningful comparison since it averages out all the seasonal trends and compared any seasonal temperature trends or annual temperature trends against the baseline average to visualize how the temperature has changed in the same regions compared to itself, and how certain regions may have experienced a more severe/normal trend. This aids in understanding which variables have more of an effect on temperature changes in the US.

For my study I have used variables such as CO2 analysis, seasonal changes, annual averages with population density changes per state to see the relationship between temperature and these economic variables that may aid in further understanding of the research question.

For the temperature change in the US states overtime I have discovered that there is a larger change in temperature increases for US states that are geographically located in the north and western region. Furthermore, I have found that using my summary statistics data and mapping over the US states grouped by the seasons, that the winter months in any given state are a lot more prone to temperature variability and have changed a lot more in temperature compared to that given state's baseline year average temperature. To make an accurate comparison overtime, I have used the mean temperature from years 1900-1950 as the "baseline year temperature", and compared this temperature to the temperatures against other years as benchmark, on a state by state basis. For example, when isolating seasonal trends per state, I have taken my data and calculated each season's average for that state from 1900-1950 using monthly observations, and then found the percent change for each observation by comparing that observations monthly state temperature by the baseline temperature for that state in that season. Then using the results for each observation in a given season, I have averaged each states percent change for the 4 seasons to map this analysis.

These is also a seasonal trend that I have uncovered, and have noticed that this trend is more prominent in northern and southern states. In the winter time, northern states have had a larger and more severe trend towards a decrease in temperature than the southern states compared to its winter baseline temperature. Also, in the spring and fall time, there is a larger temperature decrease trend in northern states compared to their relative baseline temperatures that are closer to bodies of water. This may point to the fact large bodies of water may be related to larger temperature variation in the spring and fall in neighbouring states.

For an economic interpretation, I have also included population percent changes per decade for each state and compared that against the percent temperature change in each decade per state. While I could not uncover the linear trends because I had too little data for population percentage change, I noticed that some states, while having a large population overtime, do not have large temperature changes, indicating that population may not play a large part in the states that have less temperature variability than geographic location.

Data

I have uses the "Earth's Surface Temperature Data" data set to conduct my research. This raw data comes from Berkeley Earth which is affiliated with Lawrence Berkeley National Laboratory. The data used for this paper is the "Land Monthly Average Temperature (1750-present)". I have isolated only the US state temperatures. Each observation is the monthly average temperature in degrees Celsius for that given US state in a year (indexed from 1900-2013). There are 69,613 observations in total.

I have other data sets that are combined with this data set. For example, there is a population density data set per state from the years 1910-2020. Each observation is a state in a given year from 1910 to 2020 that shows the population percentage change in that state. There are 612 observations in total. The source of this data set is the United States Census Bureau.

I also used API-web scraping to obtain a data set with CO2 emissions per state. The source of this data is the US Energy Information Administration. There are a total of 63, 063 observations. This data provides information on carbon dioxide emissions for each state in the United States from 1970 to 2020, on an annual frequency. The data set includes information on the amount of emissions produced by various sectors, such as residential and transportation, and from different types of fuels, such as coal and natural gas. The emissions are measured in million metric tons of CO2, and the data set also includes information on the units of measurement for each value. This data can be used to analyze trends in carbon dioxide emissions over time and to compare emissions across different states and sectors.

Summary Statistics

For the main temperature data set, first I grouped the states together and calculated the average temperature and standard deviation per state. Since the pandas data frame was too large to include into this file, here is the visualisation results:

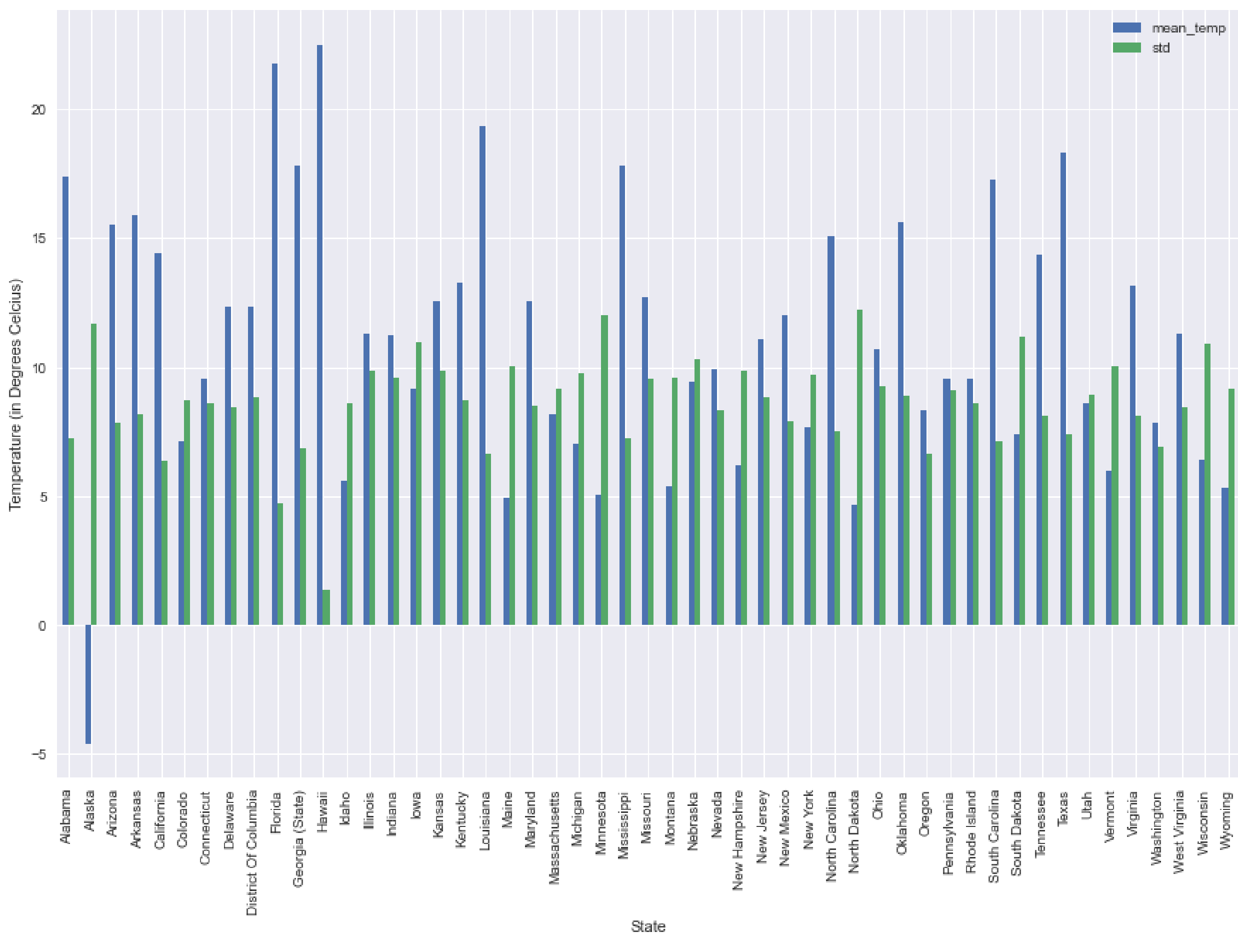


Figure 1. Average and standard deviation of temperature for each state in the US from 1900-2013

This plot displays the average temperature and standard deviation of the average temperature overtime using 1900-2013 historical monthly averages. From this we can determine how certain states may behave in terms of temperature changes over the course of our analysis. For example, we can see that certain states have a lower standard deviation than their mean, hence we may observe these states have lower temperature variability in any given year or any given season, while other states that have high standard deviation (higher than its mean temperature) may be have more variation in temperature changes.

Using this analysis, we can try to determine if these results from the graph are a byproduct of the geographic location of the state or whether there are other factors that can help predict temperature changes/variability.

Now I want to see if there is a regional pattern among states for the temp. However, geographically speaking, this is going to be a biased since certain states in certain regions are more 'hot' than others because of their location (closer/further from the equator). Therefore, the classification of states in terms of the average temperature may not be that meaningful for the research question. To solve this problem, I can take the average temperature percent change of each state overtime and see some prominent changes in certain regions better to further aid my hypothesis.

However, from this result, we can analyze the standard deviation (variation) of the average temperature of each state, and maybe this will tell us something about the variation of the temperature for each state. After all climate change can not only be defined through the overtime increase in temperature, but also higher variability in temperature overtime. Maybe there are certain states in certain regions that experience more variability in temperature changes overtime. I will also look at seasonality, since for certain states (such as Alaska or Hawaii), the temperature changes might be more variable or systematically higher/lower on average due to its location, so taking the aggregate temp. may not be the more efficient way to analyze long term trends.

So overall, for a better look that the research question, I think its more meaningful to analyze the temperature change, standard deviation trends and any other potential variables overtime, instead of getting the overall average. Here is a snippet of the data frame that shows the standard deviation and average of all four seasons for all states:

State	AvgFallPercent Change	AvgSpringPercent Change
Alabama	0.524667	1.207958
Alaska	-1.654958	-3.666642
Arizona	1.308517	1.192448
Arkansas	0.393027	1.642405
California	1.900165	0.985690

Table 1. Summary statistics for each state and its respective seasonal change from its baseline average values in that season

Since the data was too large, I left out the rest of the states and the remaining six rows. This is the summary statistics used for mapping in the next section

. The measure of percent change is used against the baseline years because it takes into account the size of temperature magnitude and it normalizes it according to the analysis of change instead of size. Using these results, in the winters, there is a lot more temperature variability compared to the other seasons. This is an interesting find, although not too surprising. One reason why it may be easier to see climate trends in states during winter months is that the temperature variations during this season tend to be larger than during other seasons, especially in northern regions of the US. This means that any changes or anomalies in temperature would be more apparent during the winter months.

However, it is also important to note that climate change impacts are not limited to changes in temperature alone, and other factors such as precipitation, extreme weather events, and sea level rise can also have significant impacts on different regions. Therefore, it would be important to analyze and consider these factors as well in order to fully understand the climate patterns and their potential impacts on different regions.

Next, for a regression analysis I want to conduct for seasonality, I want to make sure that we can model temperature change overtime as a liner trend, so the model coefficients and the regression is statistically significant. In order to determine this I use the original data set without any grouping and plotted all the state temperatures overtime. Here is the result:

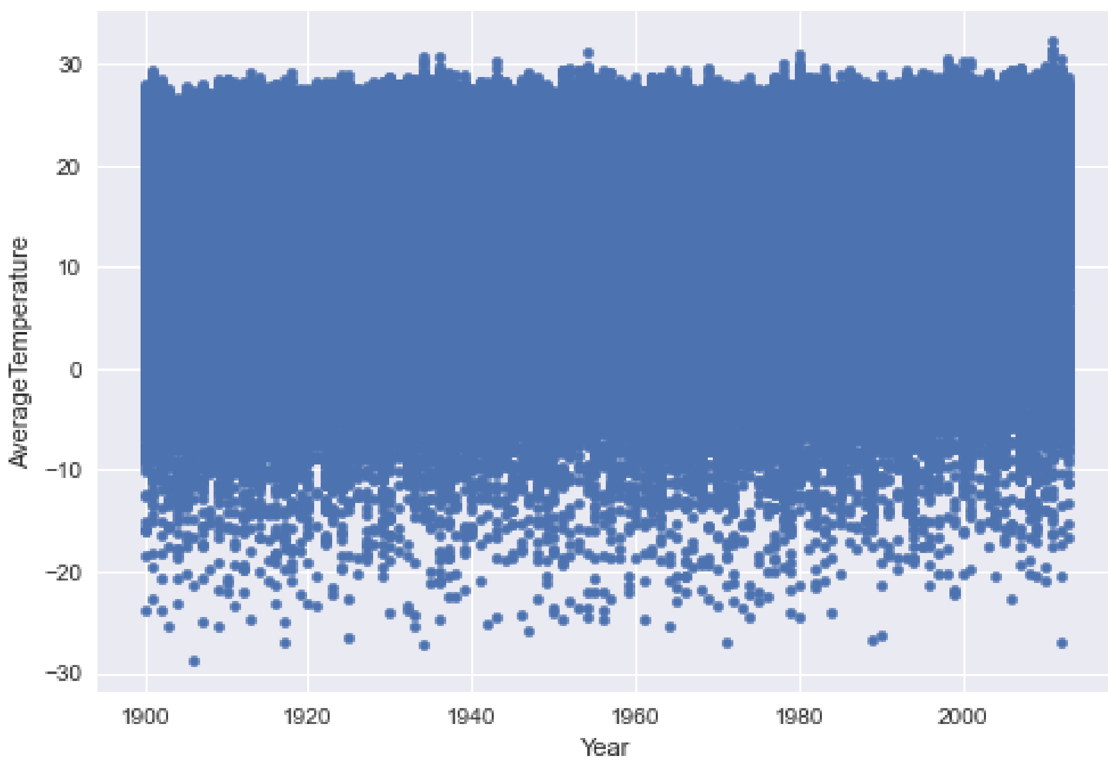


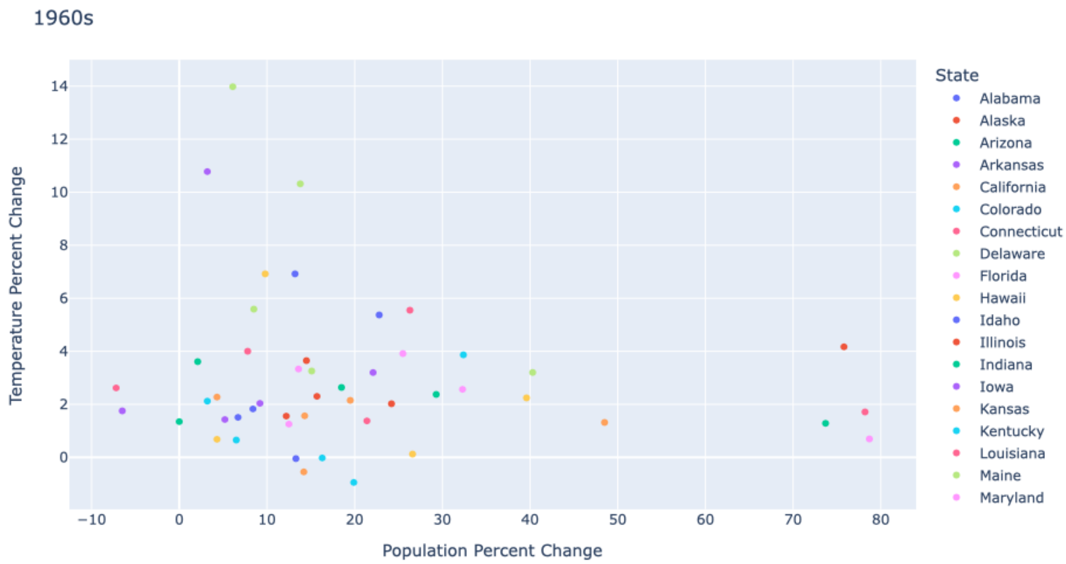
Figure 2. Average temperature for all state in the US from 1900-2013. Note there are many observations because each state is aggregated monthly, so there are 12 observations per state in each year.

Since there is a clear linear trend, and we can see it is slowly increasing overtime, a linear regression would be appropriate.

Next, for the population percentage change data set, I have too little observations for a linear regression, so I decided to do a visualization against temperature change instead, to look at some general trends. Here are the graphs for the population percentage change each and the temperature change compared to the baseline year averages per state (Note: since the population percentage change data starts from 1910, I have omitted graphs from the 1910-1950 range since these are the baseline years and we want to look at the results following these years):

The temperature percent change in the y axis and the population percent change in the x axis.

Overall, there is a general trend: We can see the much hotter states such as California, Idaho, Florida, Nevada and Arizona have a consistently higher population percent change overtime, while their temperature change overtime remains consistently closer to 0 than most other states. This overall trend is consistent with our seasonal findings (which you will see in the mapping section): The hotter states exhibit an overall LESS volatile change in temperature as compared to the northern and northeastern states, or 'cooler' states. Geographically speaking, this trend means that for mid-southern states, population changes may not be the most accurate determinant of temperature change overtime; that the change can be denoted to the geographic location of the state. Again I have to reiterate, correlation does not equal causation, so we can't make any causal inferences regarding this observation. This can merely be ONE of the many variables that effect temperature overtime.



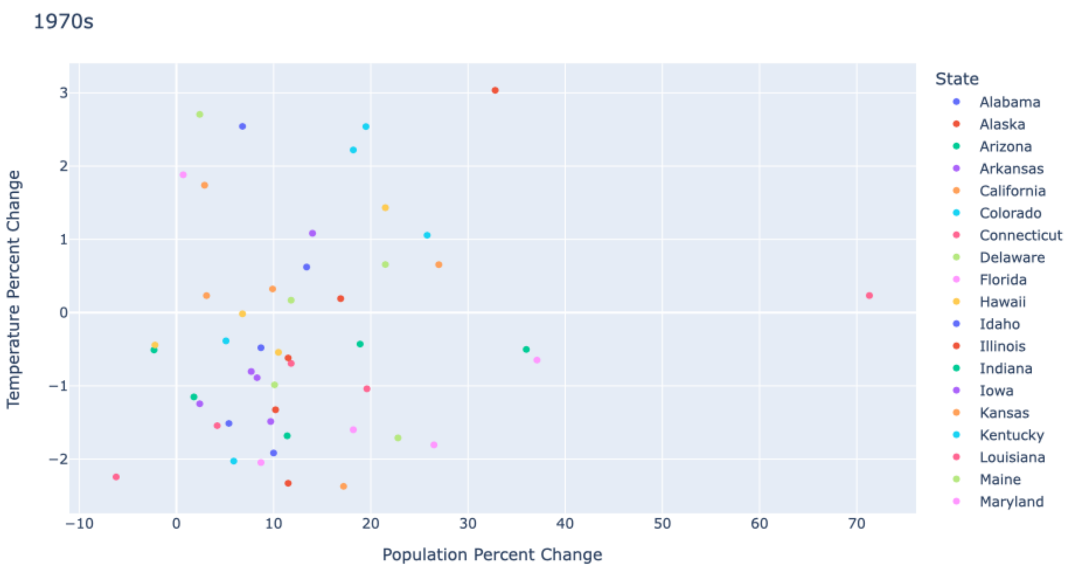


Figure 3. Population percentage change plotted against temperature percentage

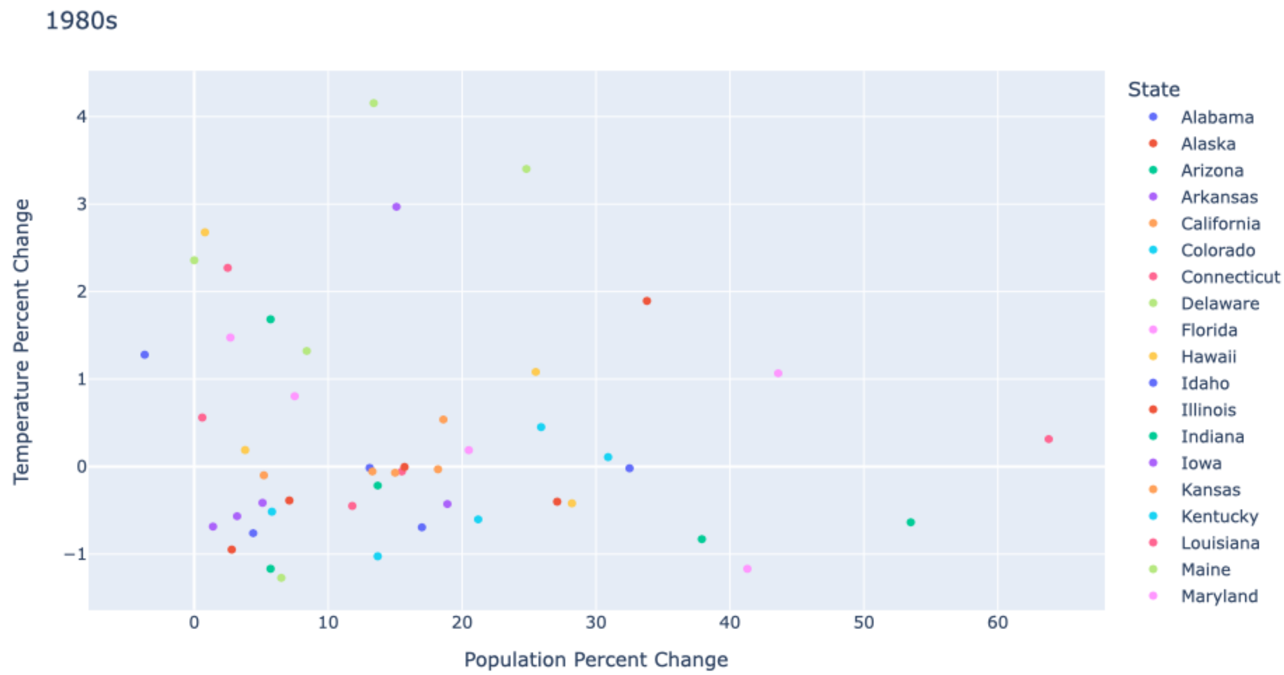
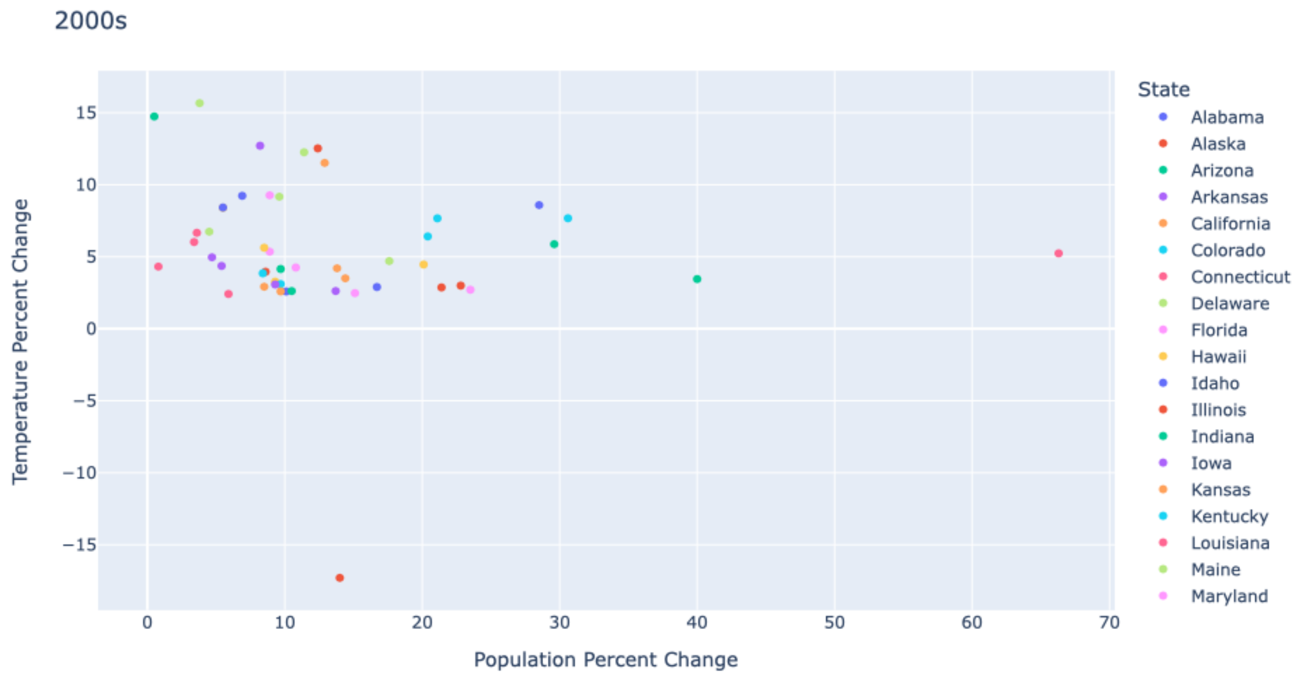


Figure 4. Population percentage change plotted against temperature percentage



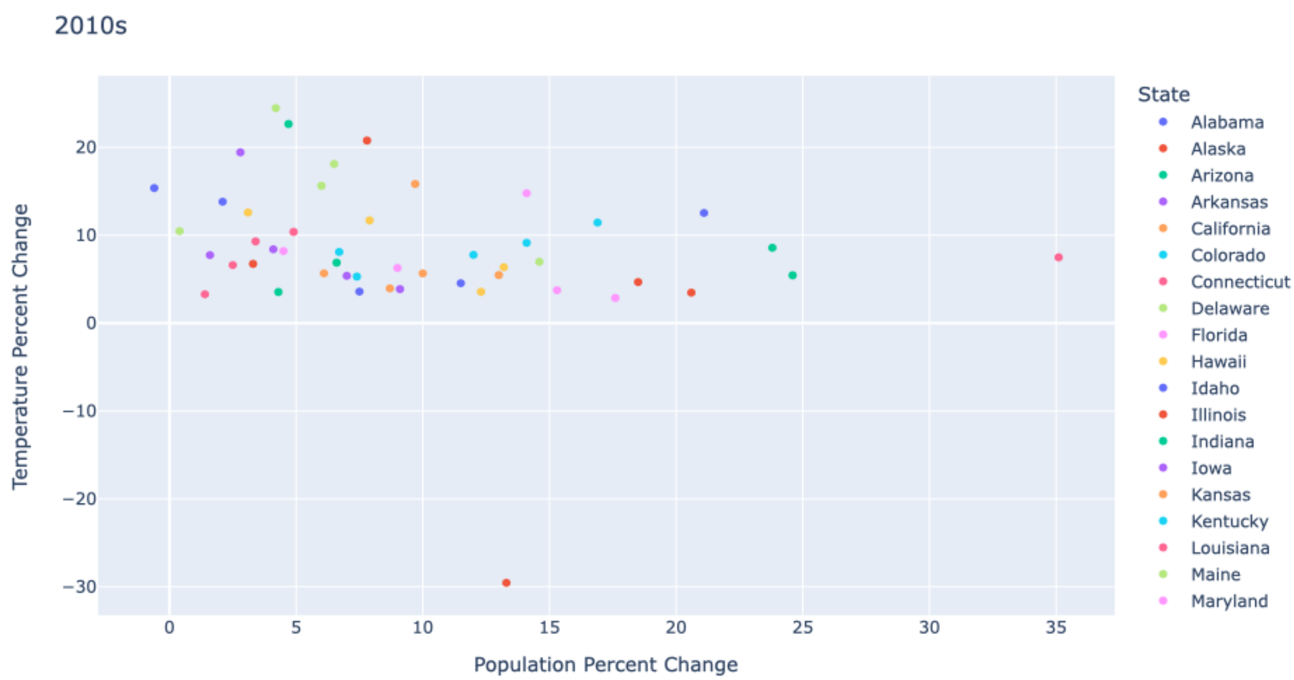


Figure 5. Population percentage change plotted against temperature percentage

There is another very significant trend. As we can see, starting the 1990's for most states (again, Alaska being an outlier due to its geographical location), the temperature change veers more towards a positive increase, and less states have a mean temperature change (compared to the baseline years) that has REDUCED from the baseline year mean. Meaning that, for each state's baseline yearly temperature (1910 average, 1920 average ... 1950 average), compared against its monthly temperature and then taking the average of those comparisons for that decade, we can see that in 1990, 2000, and 2010, the temperature change is systematically higher than its baseline.

Visually, this can be seen by the 1990, 2000, and 2010 scatter plots, where the temperature change is always positive. This is so interesting to see this visual trend because it confirms something about our data, whether its relevant to the population variable or not (which in itself, is a good indicator of whether or not this is an economically significant variable in determining climate trends). We can see from these scatter plots that overtime, the temperature change has gone from negative change to a positive change overtime, so there is definitely some form of climate warmth in the US in each state, which is especially seen from 1990 onward.

Regarding population change factor, we note that there are some patterns that may be important such as historically warmer states being less sensitive to population changes overtime.

Overall, we really can't make a strong connection using a linear regression because we don't have enough variables for the population change overtime. If we had, for example, monthly data for percent change like we did for the temperature, then we could have run those as x and y variables side by side, so this is a pretty limited conclusion, but one still worth pursuing.

Next to run regression on the CO2 emissions data-set, I have graphed the CO2 emissions by sector CO2 output (Residential, Commercial, Transportation, Total, Electric, Industrial). We can see that there are some states that fluctuate overtime but overall, the trends are mostly linear:

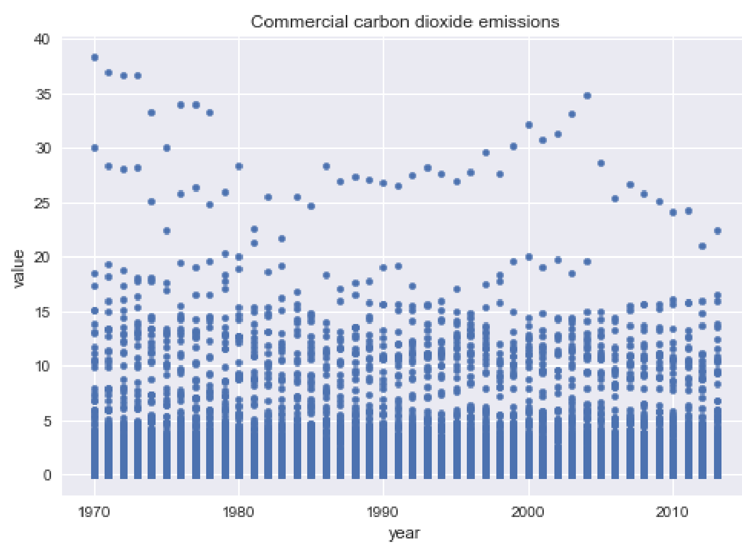


Figure 6. CO2 emissions by sector aggregated by state

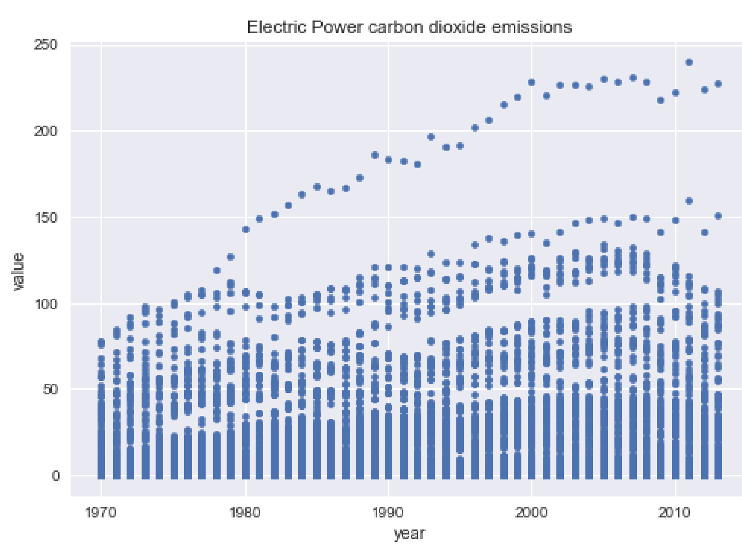


Figure 7. CO2 emissions by sector aggregated by state

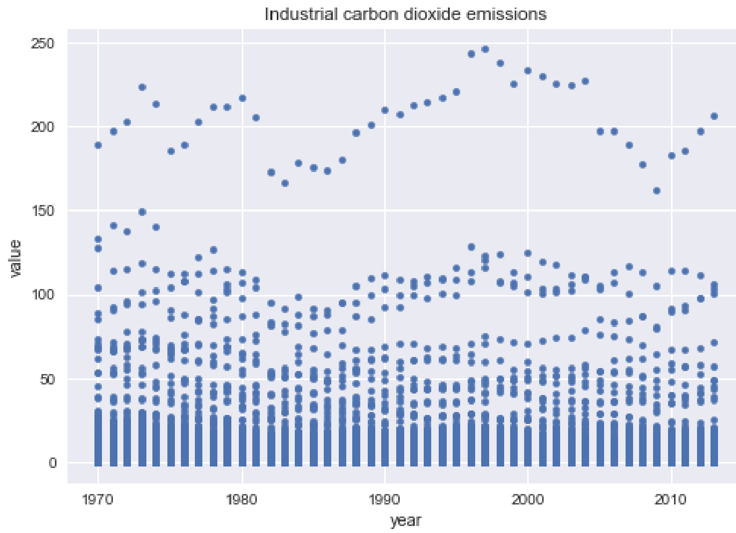


Figure 8. CO2 emissions by sector aggregated by state

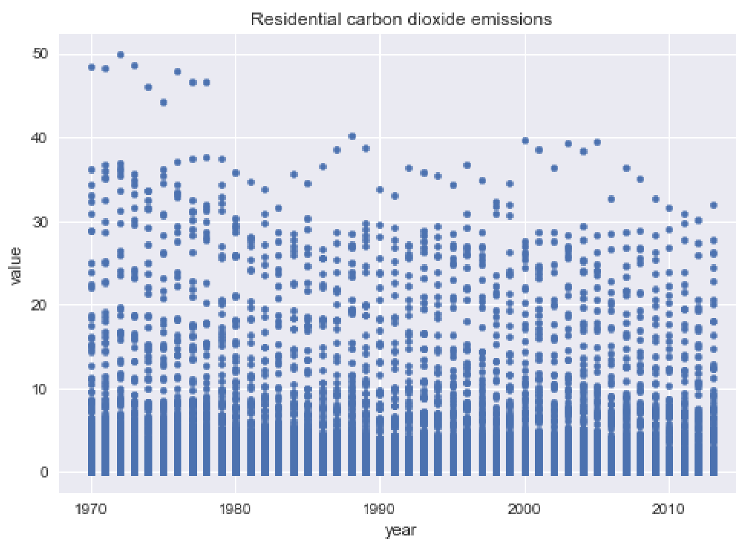


Figure 9. CO2 emissions by sector aggregated by state

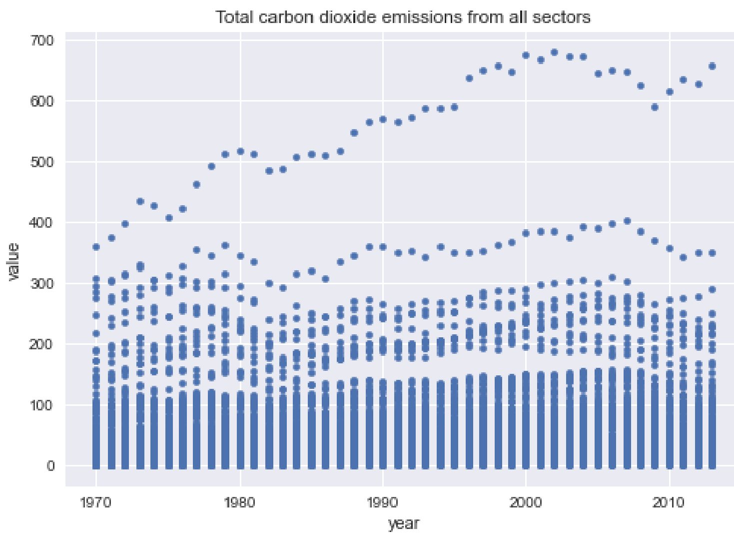


Figure 10. CO2 emissions by sector aggregated by state

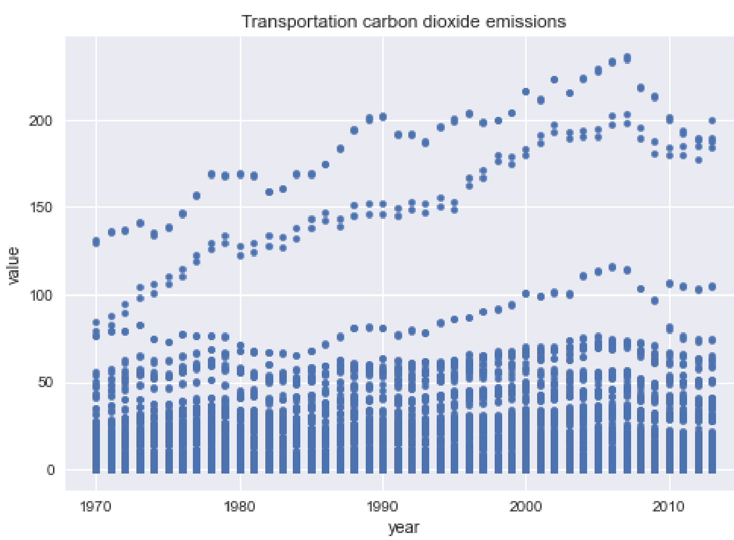


Figure 11. CO2 emissions by sector aggregated by state

Visualization

In this section we will look at some mapping results and other scatter plots that are relevant in answering our research question. First I want to point out the states have a higher standard deviation than their mean temperature, as I pointed out earlier, these states may be more prone to temperature changes because of their geographic location. Here is a graph that shows those states. They are Michigan, Massachusetts, Idaho, Nebraska, South Dakota, Colorado, Utah, Wyoming, New York, Alaska, Vermont, Montana, Iowa, New Hampshire, Maine, Wisconsin, North Dakota, Minnesota:

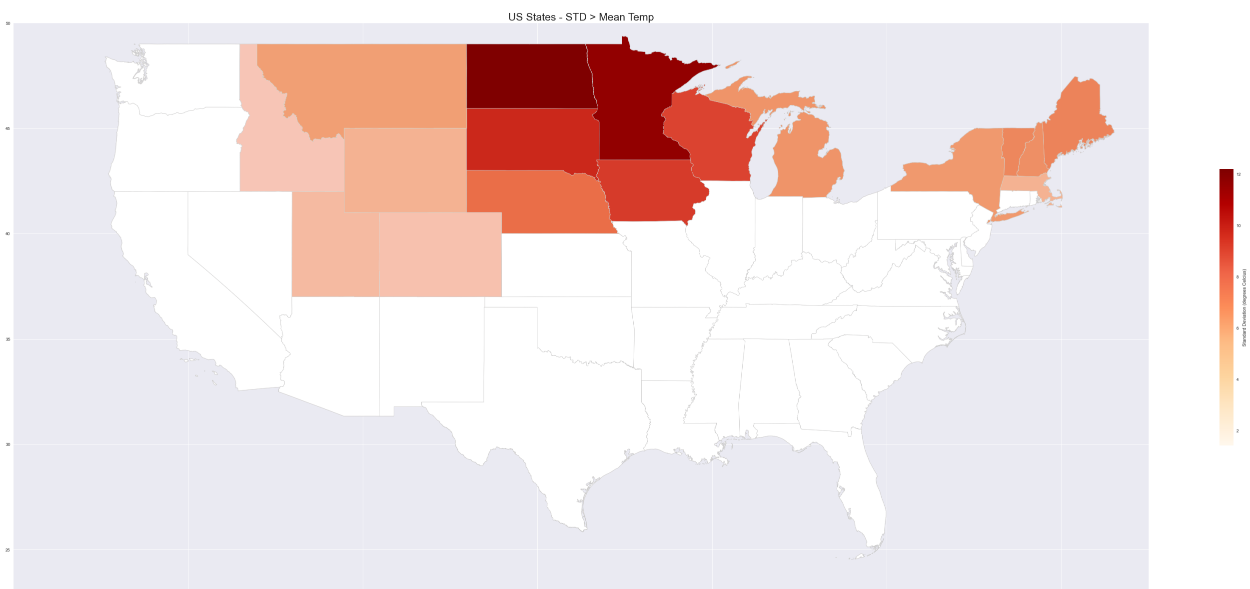


Figure 12. States that have a larger standard deviation compared to its mean temperature from 1900-2013.

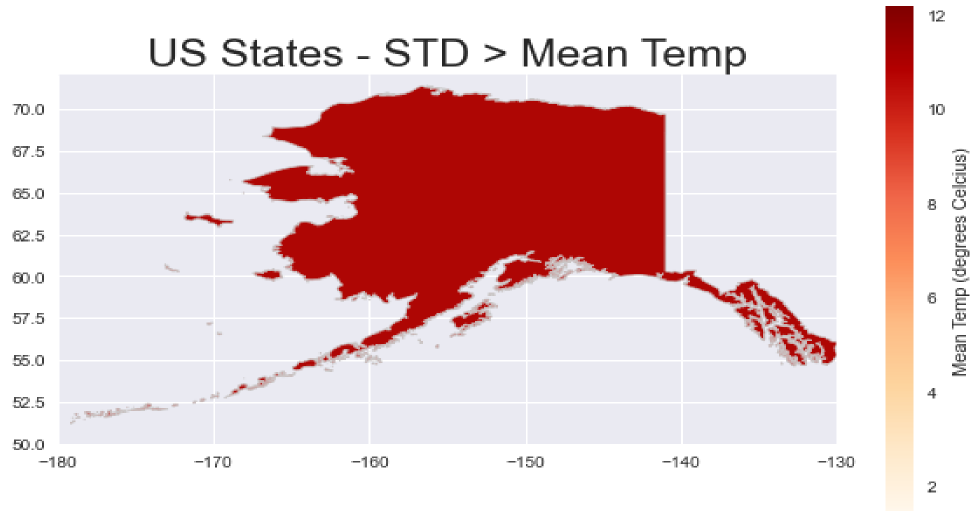


Figure 13. States that have a larger standard deviation compared to its mean temperature from 1900-2013.

Here are also the graph for the overall mean temperature of the states from 1900-2013:

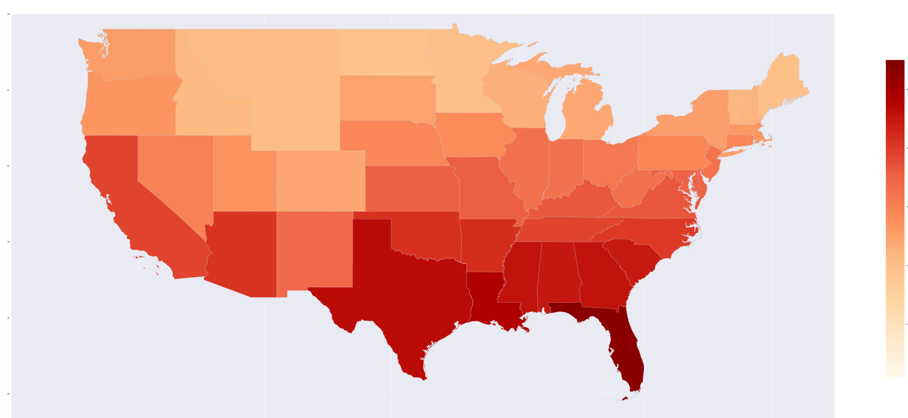


Figure 14. States aggregate mean temperature from 1900-2013.

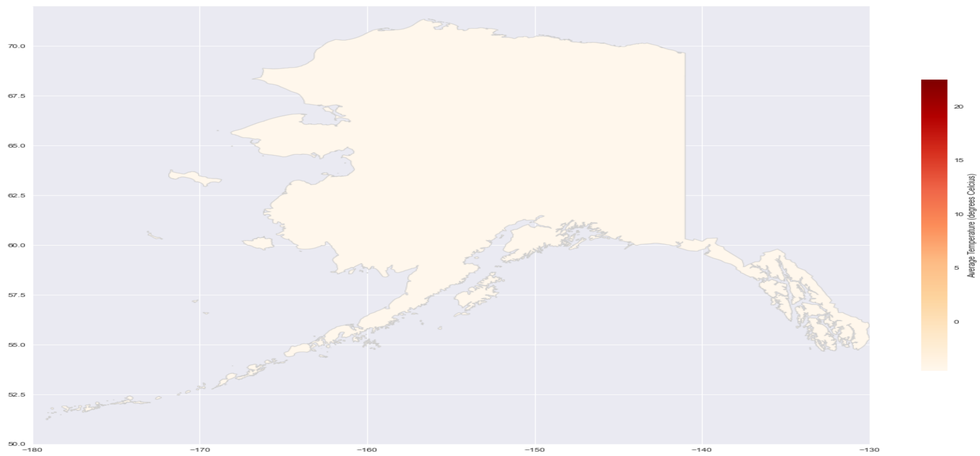


Figure 15. States aggregate mean temperature from 1900-2013.

According to our previous analysis with population, we can see that the states with lower mean temperature on average have a higher standard deviation, and vice versa. So according to these maps the northern states have higher temperature variability overtime. The hotter states exhibit a more stable temperature change overtime, and therefore, may be less prone to changes in other economic variables such as population changes. However, these graphs do not control for seasonal trends (as we can see Alaska's graph has a very high temperature), therefore, lets look at the seasonal temperature changes compared to the baseline seasonal temperatures per state. Here are some maps that control for season temperature changes. These maps are constructed using the data frame talked about earlier which uses the baseline average per season per state from 1900-1950 and finds the percentage change for each month against that baseline value and aggregates it per state.

According to these visualizations, the most drastic temperature change from the baseline years until 2013 is during the winter. It is even more interesting to see that the northern states have a temperature change that is lower than the winter baseline average for that given state. This means that compared to the baseline historical average of these states

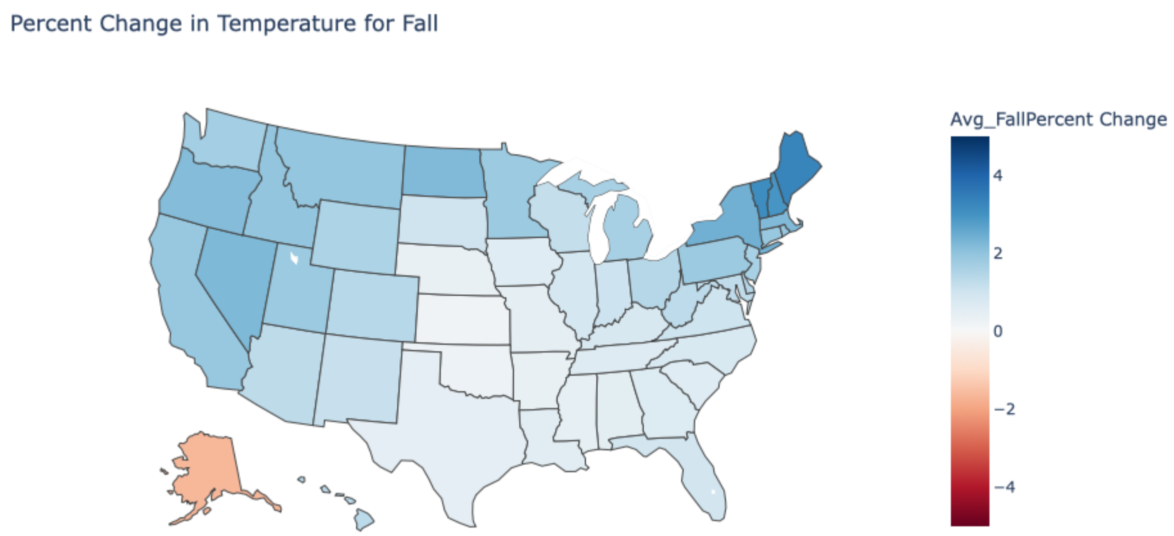


Figure 16. State temperature change per season using 1900-1950 season's as the baseline average compared against years 1900-2013.

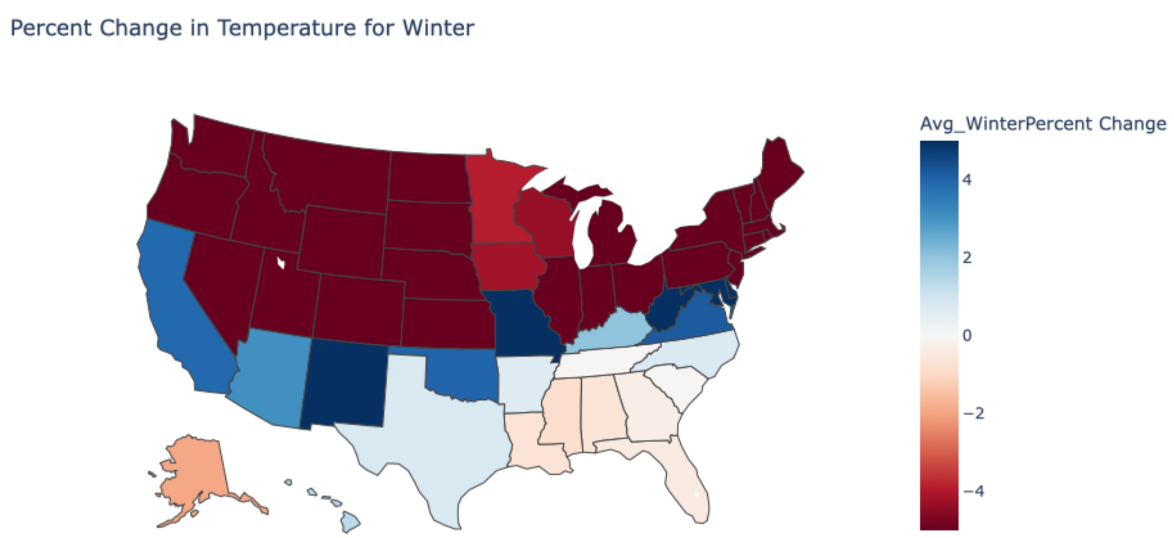


Figure 17. State temperature change per season using 1900-1950 season's as the baseline average compared against years 1900-2013.

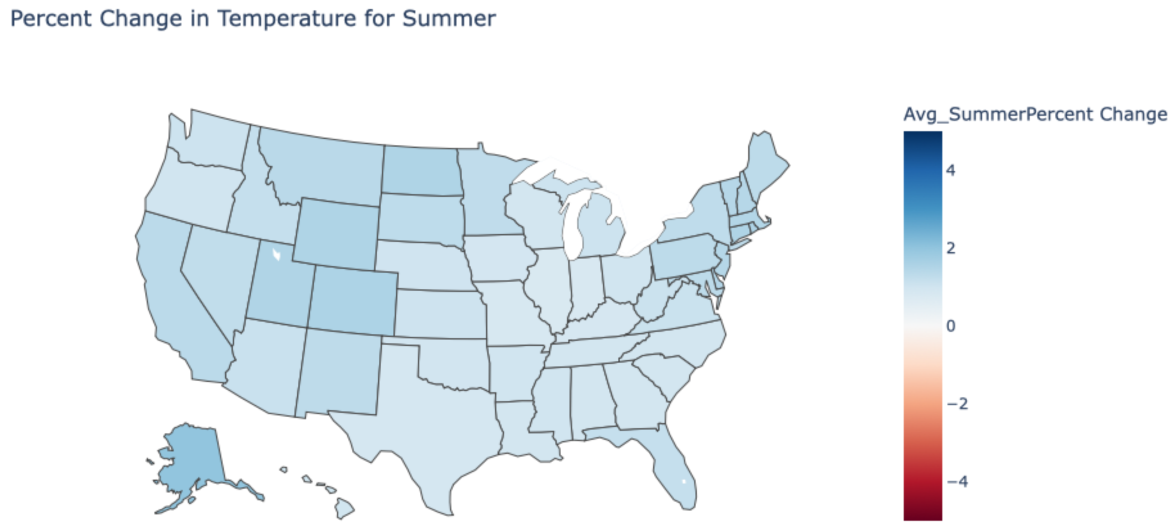


Figure 18. State temperature change per season using 1900-1950 season's as the baseline average compared against years 1900-2013.

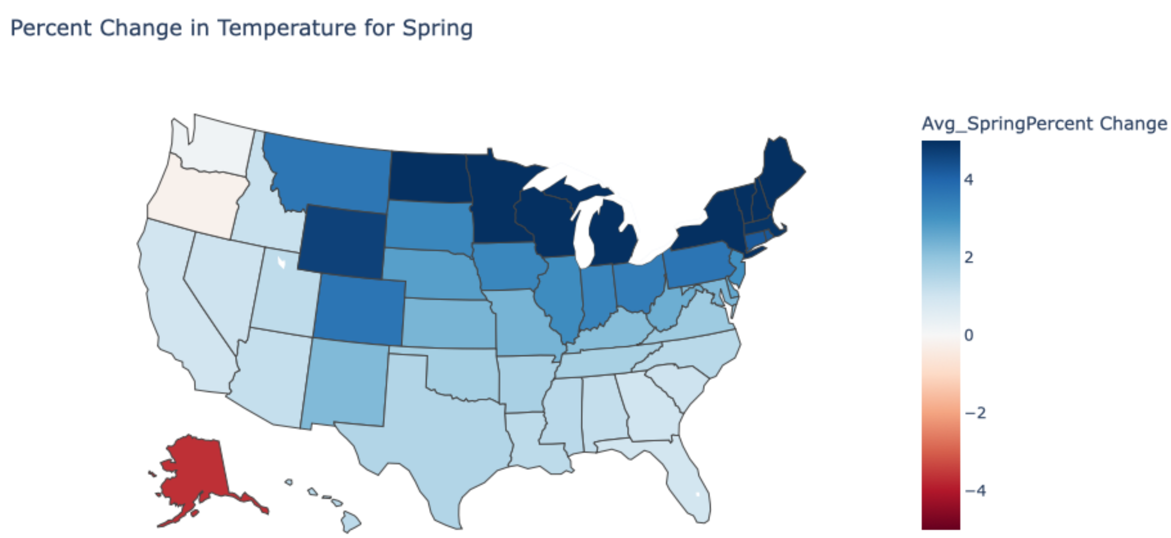


Figure 19. State temperature change per season using 1900-1950 season's as the baseline average compared against years 1900-2013.

in the winter months until 1950, the winter average temperature in the winters has decreased more on average. Some of these states are also the ones we mapped earlier that have a large standard deviation in comparison to its mean. We can see that in most seasons (not summer) the northern states have a generally more volatility in temperature. The summer months are usually close to the baseline average, not exhibiting much volatility in temperature even until 2013.

The temperature change in spring is also quite different. The northern states, again, exhibit a larger temperature change. Compared to the baseline years average temperature in the spring months, the average temperature has increased more on average, in these northern states. We can note that these states are cooler over average, and are closer to colder bodies of water (The Great Lakes and the Atlantic Ocean) which may get cooler/warmer during the spring due to a drastic shift in temperature. In these northern states, the changing of season's from cold to hot can take longer, therefore, the shift in temperature of the bodies of water and the states may be more variable, reflecting that.

So as we can see, there are regions that exhibit different changes in climate overtime, and in the next section I will fortify these justifications using regression analysis.

Regression Results

For the first regression, I wanted to isolate seasonal trends and well as annual changes. Therefore, the first regression I ran used the original data set and created four dummy variables for seasons. In this regression, there are a total of five independent variables and the dependent variable is the average temperature in that month. I ran a total of 50 regressions, grouping by state. Since the regressions results are too long I will add a few here and talk about the rest.

The model can be introduced like this:

$$AverageTemperature_i = \beta_0 + \beta_1 year_i + \beta_2 summer_i + \beta_3 spring_i + \beta_4 winter_i + \beta_5 fall_i + u_i$$

where:

- β_0 is the intercept of the linear trend line on the y-axis

- β_1 is the year variable explaining how much temperature increases by (in Celsius) annually, on average in that state

- $\beta_{2..5}$ are the dummy variables that represent the average temperature that year in that state, or 0 if it is not that season. So, on average, how much the temperature changes in Celsius during *that season* compared to the baseline years' average temperature in *that season*, taking into account yearly changes and controlling for other seasons.

- u_i is a random error term (deviations of observations from the linear trend due to factors not included in the model) .

	Dependent variable:AverageTemp	
	Alabama	Alaska
	(1)	(2)
Fall	1.482	-5.665***
	(1.106)	(1.833)
Spring	0.810	-6.630***
	(1.107)	(1.835)
Summer	9.742***	9.384***
	(1.107)	(1.835)
Winter	-8.260***	-19.416***
	(1.106)	(1.834)
Year	0.006**	0.012**
	(0.003)	(0.005)
const	3.775	-22.328***
	(4.379)	(7.260)
Observations	1,365	1,364
R^2	0.780	0.768
Adjusted R^2	0.779	0.768
Residual Std. Error	3.394(df = 1360)	5.621(df = 1359)
F Statistic	1205.077*** (df = 4.0; 1360.0)	1127.383*** (df = 4.0; 1359.0)
Note:		*p<0.1; **p<0.05; ***p<0.01

Here, each model corresponds to a regression run for a state. I want to isolate the effects of seasonality for any given state, and see how the results of the states differ, i.e. which state may not have any or all statistically significant dummy variables, and why this might be the case.

I will do an interpretation of the results for Alaska for example, and then point out some general trends:

The multiple regression model shows that the season variables (Winter, Fall, Spring, and Summer) and the year variable have a significant impact on the average temperature in Alaska. The R-squared value of 0.768 indicates that the model explains 76.8 percent of the variance in the dependent variable (Average Temperature). This means that the model is a reasonably good fit for the data. The coefficients for the season variables show how much the average temperature in Alaska changes for each season, controlling for the other seasons, and year. The Winter coefficient of 19.416 indicates that, on average, after controlling for other seasonal trends and yearly temperature changes, the change in temperature in Alaska during winter is 19.4 degrees Celsius lower than the average temperature. Similarly, the Fall coefficient of 5.655 indicates that the temperature during the fall season is 5.655 degrees Celsius lower than the average temperature. The coefficient for the Year variable is 0.012, which means that for every one-year increase in the Year variable, the temperature in Alaska increases by an average of 0.012 degrees Celsius, controlling for seasonality.

The p-values for all of the seasonal coefficients is significant at 1 percent, the p-value for the annual coefficient is significant at 5 percent. which means that they are statistically significant at the 95 percent confidence level. This suggests that the explanatory variables are significant predictors of the average temperature in Alaska overtime.

Intuitively the coefficients make sense. The other seasons (Winter, Fall, and Spring) seem to have a negative impact on the average temperature in Alaska, while the Summer season has a positive impact (can be seen by the signs of the coefficients). This could be because Alaska is a colder state and experiences relatively mild summers, so the lower and more drastic temperatures in the other seasons has a greater impact on the temperature changes overtime. It could also be due to other factors like precipitation, cloud cover, and wind patterns, which can vary between seasons and affect the average temperature differently.

To assess my regression results, I will point out some trends that can be generalized among all the states. It's worth noting that the coefficient for the year variable is statistically significant for all states, indicating that temperature has been increasing over time, on average, across all states, controlling for seasonality. Even when we take into account seasonal temperature changes in our model, time plays a significant role on predicting temperature in each state.

Most states have a p-value that is not statistically significant for the spring or fall month dummies.

Overall, most states have either the winter or summer dummy as a statistically significant predictor of the average temperature. It makes sense intuitively that the winter and summer dummies are more likely to be significant predictors of average temperature compared to spring and fall dummies. Winter and summer are the two most extreme seasons in terms of temperature. Therefore, it is more likely that the winter and summer dummies will have a stronger relationship with the average temperature compared to the more mild seasons of spring and fall. Additionally, the winter and summer seasons tend to have more distinct and consistent weather patterns, which can also contribute to their significance as predictors.

Next, I will also do a regression for each season. This aggregates the states results together so we only get to look at the seasonal changes overall. In this regression I will have the independent variables: Baseline Temperature for that season in that given year in that state, Temperature Change against the baseline in that month, and Year. The dependent variable will be Average Temperature in that given month in that season.

My reasoning for running this regression is that I want to isolate seasonal temperature effects by 1) first, controlling for that season's average temperature as a whole and then 2) look at the temperature change (fluctuations from that average) from that baseline average and in the end, looking that how significant these two might be in determining that season's overall temperature. This regression will make clear if the temperature fluctuations are a good indicator of average temperature - meaning that they are somewhat repetitive and have a clear pattern in any given state on a seasonal basis when regressed against average temperature.

The model is as such:

$$AverageTemperature_i = \beta_0 + \beta_1 year_i + \beta_2 BaselineTemp_i + \beta_3 TemperatureChange_i + u_i$$

where:

- β_0 is the intercept of the linear trend line on the y-axis

- β_1 is the year variable explaining how much temperature increases by (in Celsius) annually, on average in that season

- β_2 is the baseline average temperature in degrees Celsius for that season in a given year in a given state, aggregated for each state

- β_3 is the temperature percent change in that given season for a given month for a certain state, which was calculated by the formula (average temp monthly temp in a state in a given year for a given season - baseline average temp for that season in that year for that state / baseline average temp for that season in that year for that state), in percent.

- u_i is a random error term (deviations of observations from the linear trend due to factors not included in the model).

	Dependent variable:			
	Fall	Spring	Summer	Winter
BaselineAvgTemperature	1.004***	1.008***	1.012***	0.981***
	(0.005)	(0.004)	(0.001)	(0.003)
TemperatureChange	0.071***	0.054***	0.195***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Year	0.003***	0.004***	0.001***	0.011***
	(0.001)	(0.001)	(0.000)	(0.001)
const	-5.181***	-6.977***	-1.558***	-21.012***
	(1.533)	(1.272)	(0.137)	(1.192)
Observations	17,338	17,442	17,442	17,391
R^2	0.787	0.842	0.994	0.876
Adjusted R^2	0.787	0.842	0.994	0.876
Residual Std. Error	3.370	2.820	0.297	2.637
F Statistic	21299.869***	31038.304***	991673.491***	40867.928***
Note:			*p<0.1; **p<0.05; ***p<0.01	

As we can observe, most of the coefficients are statistically significant at the 99 percent confidence level. This means that for each season, the baseline temperature is a strong predictor of average temperature. This is not a surprising result since the baseline is calculated from the average temperature aggregations. The main reason for having this variable in this regression is to interpret the temperature change variable. As we can see, temperature change is statistically significant for all seasons except winter. So, for the interpretation, we can say that the overall fluctuations around a given mean temperature (baseline value) for summer, spring and fall are still statistically significant predictors of temperature activity. Meaning that the fluctuations need not be random, and there can be a somewhat predictability around the magnitude of these changes around the baseline values. However, for winter, the temperature fluctuations are not significant predictors of temperature at all. The coefficient is 0, meaning that these fluctuations do not contribute to finding the average temperature overall, while baseline values are still statistically significant. Intuitively, this means that winter fluctuations are too volatile to be good predictors of temperature. Alternatively, a downside of this regression is that we cannot see this breakdown on a state-by-state basis (the regressing would be too large and we would have to create dummies for each state), so we don't know if there may be a few states that are skewing the results (e.g. Alaska or North Dakota).

Next, for the other two regressions, I will use the CO2 emissions data.

Here we can't group by season since the CO2 data is on an annual frequency, so after merging our temperature data set with the CO2, I created 6 dummy variables (for the next regression) for each sector CO2 output (residential, commercial etc) to isolate these effects.

The first regression will simply be on a per state basis and we will use the CO2 emission total values along with the year variable to predict the temperature to see how if CO2 emissions can predict temperatures is any given state, controlling for annual increases or decreases.

Here is the model:

$$AverageTemperature_i = \beta_0 + \beta_1 year_i + \beta_2 CO2emissions_i + u_i$$

where:

- β_0 is the intercept of the linear trend line on the y-axis - β_1 is the year variable explaining how much temperature increases by (in Celsius) annually, on average in that state - β_2 is the CO2 emission value in that state overall. - u_i is a random error term (deviations of observations from the linear trend due to factors not included in the model). Again the results are too large so I will only show the first two:

	Dependent variable:	
	Alabama	Alaska
const	-34.821***	-87.222***
	(2.319)	(4.658)
value	-0.000	0.002
	(0.001)	(0.004)
year	0.026***	0.042***
	(0.001)	(0.002)
Observations	1,056	1,056
R^2	0.328	0.241
Adjusted R^2	0.326	0.240
Residual Std. Error	0.479	0.947
F Statistic	256.500***	167.310***
Note:	*p<0.1; **p<0.05; ***p<0.01	

So, the attempt to isolate state level trends were insignificant since the CO2 emission coefficients are economically and statistically insignificant for each state. This may because we are aggregating all of the CO2 sector values together instead of picking and choosing which ones may be more important. However, I think the bigger picture may be that CO2 emissions may have become more prominent in later years, and therefore, regressing them against temperature for earlier years on all sectors may not be that meaningful. Instead, let's pick and chose more important variable to group by.

In this next model, we can isolate yearly values, since in the previous regression, the only significant variables was the year variables. Therefore, there must be a correlation between the year variable, and CO2 emissions as well as temperature changes. In this next regression we can also control for the different types of sectors by creating dummy variables, to see which sectors may be more/less significant overtime. Note that we don't just use the total emissions sector in this model since we want to find out which sector may produce the most statistically significant results overtime. In this model, the y variables is the average temperature in a given year from 1970-2013.

Here is this model:

$AverageTemperature_i = \beta_0 + \beta_1 CO2emissions_i + \beta_2 ResidentialCO2_i + +\beta_3 IndustrialCO2_i + \beta_4 ElectricCO2_i + \beta_5 TransportationCO2_i + \beta_6 CommercialCO2_i + iu_i$

where:

- β_0 is the intercept of the linear trend line on the y-axis

- β_1 is the CO2 emissions in total for that given year in million metric tons of CO2

- $\beta_{2 \ldots 6}$ are the CO2 emissions dummy variables per sector type overtime for that given year in million metric tons of CO2

- u_i is a random error term (deviations of observations from the linear trend due to factors not included in the model)

Since the data frame goes from 1970-2013 per year, I will show the first three output regressions:

	Dependent variable:AvgTemp		
	1970	1971	1972
Commercial carbon dioxide emissions	0.838	0.935	0.907
	(0.548)	(0.568)	(0.581)
Electric Power carbon dioxide emissions	0.670	0.743	0.712
	(0.534)	(0.554)	(0.566)
Industrial carbon dioxide emissions	0.613	0.699	0.682
	(0.530)	(0.551)	(0.564)
Residential carbon dioxide emissions	0.792	0.884	0.859
	(0.544)	(0.564)	(0.577)
Transportation carbon dioxide emissions	0.657	0.726	0.701
	(0.533)	(0.553)	(0.565)
const	9.789***	9.727***	9.458***
	(0.419)	(0.435)	(0.445)
value	0.021***	0.023***	0.021***
	(0.005)	(0.005)	(0.005)
Observations	1,176	1,176	1,176
R^2	0.015	0.017	0.015
Adjusted R^2	0.010	0.012	0.010
Residual Std. Error	5.046	5.237	5.346
F Statistic	2.910***	3.375***	2.984***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Here we can see that all of the values for the CO2 emissions are statistically significant at the 99 percent confidence level for all years until 2013. Meaning that, when grouping by year, CO2 emissions are strong predictors of annual average temperature. However, we must note that theR^2 value is very low, meaning that model does not explain most of variance surrounding the average temperature, but its still significant. This could indicate that factors other than CO2 emissions are more relevant in predicting temperature overtime. The model however, is still statistically significant and therefore can be used for further analysis. All of the coefficients are positive in this model, meaning that CO_2 emissions and yearly temperature are positively correlated, which is not a big surprise in this case. We can also note that since this data was annual instead of monthly, a lot of valuable observations were aggregated, therefore, the model may lost some precision.

When breaking down the sector analysis, we can see that Industrial, Residential and Commercial CO_2 emissions are the most statistically significant, and therefore, contribute most to the temperature changes overtime. Also, the coefficients for these three have an upward trend, meaning that they contribute more overtime to CO_2 emissions, and therefore, to the changes in temperature. For an interpretation for the year 1970, we can say that, after controlling for the different types of sector emissions, a 0.021 degree Celsius increase in temperature can be observed if we increase overall CO_2 emissions in 1 million metric tons of CO_2 in the year 1970 for all the states in the US.

So, from these findings we can see that there are other variables that may be able to predict how temperature may change overtime. CO_2 emissions are slow acting and for this very small data set, even though the R^2 is very small, we can still make inferences on how changing variables can effect our results. We can also note that the low R^2 result can come from the fact that we look at each year's CO_2 independently, and therefore, the growth pattern amongst the CO_2 and temperature will not have that much of an impact on each single regression, since we are not looking at annual changes anymore.

Conclusion

While there are many factors that can contribute to temperature changes in individual states, there are some economic indicators that may help explain why certain states have experienced more dramatic temperature changes over time than others. Here are a few possibilities that we have uncovered through our findings:

Sector based CO2 activity: Certain sectors such as residential and industrial sectors play a more important role in determining how a state's temperature may behave. Overtime, sector based CO2 activities can become better determinants of predicting state temperature. Intuitively, this makes sense, since larger industrial activity may point to higher CO2 emissions and larger population changes may cause larger residential CO2 emissions.

We have also noted that geographically, northern states exhibit more volatile temperature changes overtime, no matter what the season is, and southern states (or the "hotter" states) are less effected by economic variables such as population changes overtime, due to their systematically high temperatures.

Overtime policymakers should take into account these economic findings to make decisions that effect state level output activity caused by more industrialization and immigration policies to assess how climate change can be effected by population changes.

Bibliography

[1] Colacito, Riccardo, Bridget Hoffmann, and Toan Phan. *"Temperature and growth: A panel analysis of the United States."* , Journal of Money, Credit and Banking 51.2-3 (2019): 313-368.

[2] Karl, Thomas R., and William E. Riebsame. *"The identification of 10-to 20-year temperature and precipitation fluctuations in the contiguous United States."* , Journal of Applied Meteorology and Climatology 23.6 (1984): 950-966.

[3] *State carbon dioxide emissions data - U.S. Energy Information Administration (EIA). State Carbon Dioxide Emissions Data - U.S. Energy Information Administration (EIA)*, (n.d.). Retrieved April 16, 2023, from https://www.eia.gov/environment/emissions/state/