Climate Change and Severe Weather Events

Authors
Jared Jacobowitz and Joshua Yoon

16 December 2021
SS340 - Professor Loujaina Abdelwahed

Contents

List of Tables

List of Figures

Introduction and Motivation

Due to emissions produced by human activities, the average global temperature has been rising at an alarming rate (NASA, 2021). According to Lindsey and Dahlman, 2021, the temperature-rise trend from 1880 - 1900 was around $0.1°C$ per decade, but since then, that rate has increased to $0.2°C$ per decade. To understand the consequences of the rising temperature trend, the causal effect of temperature on the number of FEMA (Federal Emergency Management Agency) declared disasters (as defined by FEMA, 2021a) was investigated. To narrow down the data being analyzed, only temperature (used as a proxy for climate change) and FEMA-disaster data pertaining to the contiguous United States from 1953 - 2019 was analyzed. From this report's analysis, it was expected to see an increase in the number of FEMA-declared, weather-related disasters as a result of temperature rises.

This analysis is important because it shows that climate change policies should be enacted to prevent an increase in the number of severe weather events that occur in the contiguous US. Reducing the number of disasters would reduce the number of disaster-related deaths (which could be a further analysis performed) and reduce spending needed in response to severe weather disasters.

Data

Data Sourcing

FEMA severe disaster declaration data was downloaded from the FEMA dataset website (FEMA, 2021b). Contiguous US county panel temperature data was downloaded from GitHub and The Washington Post, 2021, a GitHub repository which attributed the data to The Washington Post. The datasets spanned different year ranges, so they were constrained to the common range of 1953 to 2019.

The datasets used FIPS (Federal Information Processing Standard) codes to identify the locations. FIPS codes are a US code used to uniquely identify the state and county locations. These codes were decoded using the USDA, n.d. site.

Note that the FEMA dataset has inherent political bias as disasters are more likely to be declared if the governer is of the same party as the one in power at the Capital. Due to time constraints this bias was not accounted for in the analysis.

Disaster Research

The FEMA disaster dataset contained disaster declarations for a large range of weather-related disasters. Only disasters that were believed to increase in number and severity as a result of temperature rises, and therefore climate change, were used. The disasters used in the analysis were dam/levee breaks (Kiprop, 2018), droughts (Center for Climate and Energy Solutions, 2021; Tschumi & Zscheischler, 2019), typhoons (Michener et al., 1997), floods (Wasko et al., 2021), hurricanes (Michener et al., 1997), severe storms (Michener et al., 1997; Wasko et al., 2021), fires (Flannigan et al., 2006), mud/landslides (Banholzer et al., 2014), fishing losses, which were mainly caused by hurricanes or toxic algae blooms

(Environmental Protection Agency, n.d.), and coastal storms (Michener et al., 1997). The weather-related disasters not used in this analysis because research showed their frequency or severity did not increase with temperature were tornadoes (Brooks, 2013), severe ice/snow storms (Irland, 2000; Karl & Trenberth, 2003; Fassnacht et al., 2016), and earthquakes (NASA & Buis, 2019).

Combining the Data

The temperature and disaster datasets were combined in Python by going through each FEMA disaster declaration and incrementing the disaster count in a combined temperature-disaster dataset. Few of the disasters in the dataset were only declared as county-wide, most were declared as statewide. When a disaster was declared as statewide, all counties in that state were incremented.

Data Summary

The datasets were combined based on location and time. After combining the datasets, the combined panel dataset CSV file contained 6 columns of data about the location, time, temperature, and disaster count for each observation. These columns, and their labels, which will be referred to for the rest of this report, are outlined in Table 1.

Table 1: Analyzed dataset column label descriptions.

Title	**Description**
`fips`	Special code that designates a US county
`year`	Year from 1953 - 2019
`tempc`	Average county temperature in Celsius
`county`	US county
`state`	US state (2 letter abbreviation)
`disasters`	Statewide disaster count during the designated year

Table 2 is a summary statistics table for the temperature, `tempc`, and disaster count, `disasters`, of the entire dataset. There were a total of 208,034 observations, corresponding to an observation for each of the 67 years for each of the 3,105 counties, except for Hamblen, Tennessee which was missing an observation for 1981. The tabulated `mean` column shows the `mean` of each variable, describing that variable's central tendency. The standard deviation column, `std`, quantifies the spread of the variable's data, where a higher `std` means it is more spread out. The `min` and `max` columns show the minimum and maximum values of that variable, indicating the range of values that variable could take.

Table 2: Summary statistics of the entire dataset.

	`mean`	`std`	`min`	`max`
`tempc`	12.34	4.60	-0.26	26.01
`disasters`	0.40	2.31	0.00	55.00

Because the dataset is a panel dataset, a summary statistics table of the entire dataset does not properly show variation in either location or time (although Table 5 in Appendix section *Summary Data* shows the summary data in 5-year increments). To show the variation, a table of location and time values could be used, but that contains a large amount of data, making it difficult to understand. One way to visualize the data is through plots. Figure 1 shows the US average yearly temperature vs time data. The temperature vs time plot shows an increase in temperature over time, and will be discussed more in the *Temperature Rise* and *Temperature Regression Results* sections. Figure 2a shows the average yearly disasters, averaged per state, over time. Figure 2b shows the same graph with the two largest values, from 1996 and 1998, to remove to show a better y-axis scale. Those points were higher because, as seen in the choropleth maps (maps that shade areas by value) shown in Figures 3a and 3b, those two years had a large number of statewide disasters declared for Texas and Florida, two states with a large number of counties. (In 1996, Texas had 104 fire disaster declarations, and in 1998 had 31 fire and 262 severe storm declarations (all of which are county counts). In 1998, Florida had 80 fire, 57 hurricane, and 58 severe storm declarations.)

More choropleth maps are shown in Figures 4 and 5, visualizing the temperature and disasters over time, respectively (Note: the disaster plot was converted to a logarithm to better capture the range of values).

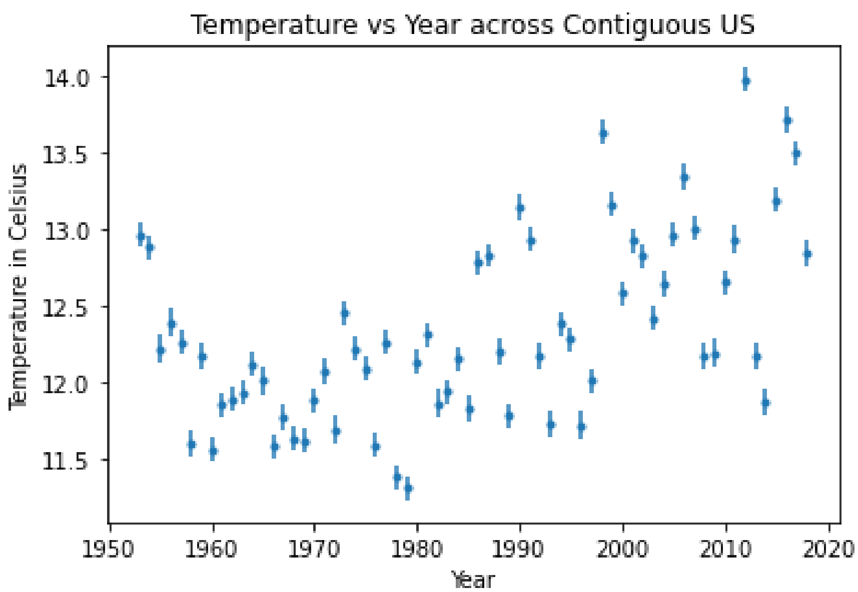
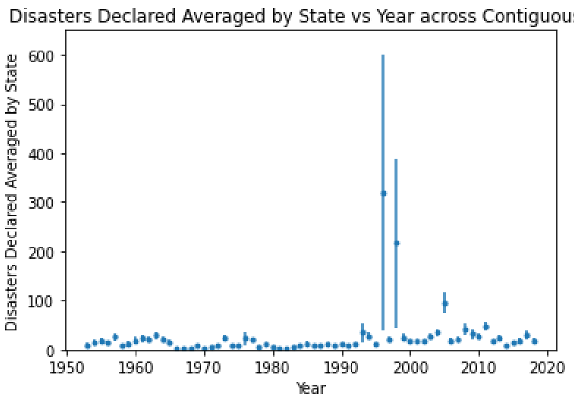
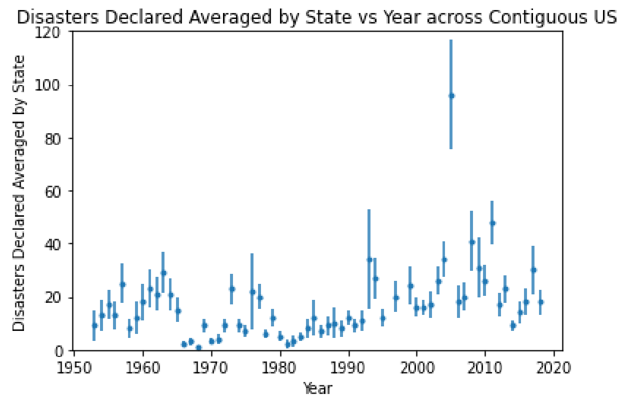


Figure 1: US average yearly temperature with standard error bars.

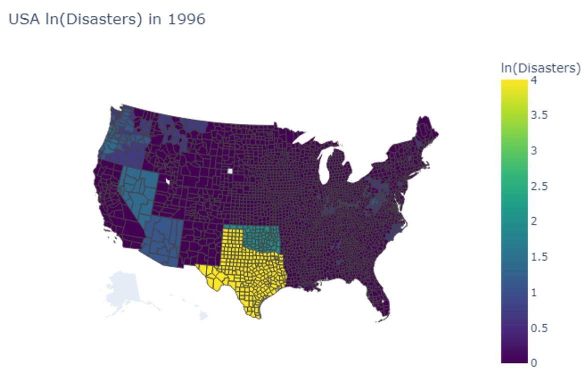


(a) All Data

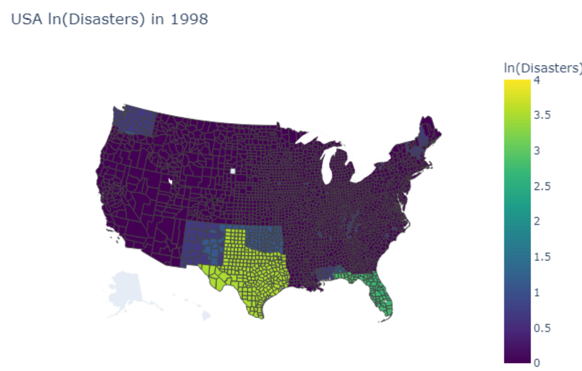


(b) Removed 1996 and 1998

Figure 2: (a) US average yearly per-state disasters for the entire dataset. (b) US average yearly per-state disasters with the larger 1996 and 1998 removed to have a clearer y-axis. *Note*: All error bars are standard errors.



(a) 1996 Choropleth Disaster Map



(b) 1998 Choropleth Disaster Map

Figure 3: Choropleth disaster maps for (a) 1996 and (b) 1998. Note the high number of county declarations in Texas in both 1996 and 1998 and the higher number in Florida in 1998.

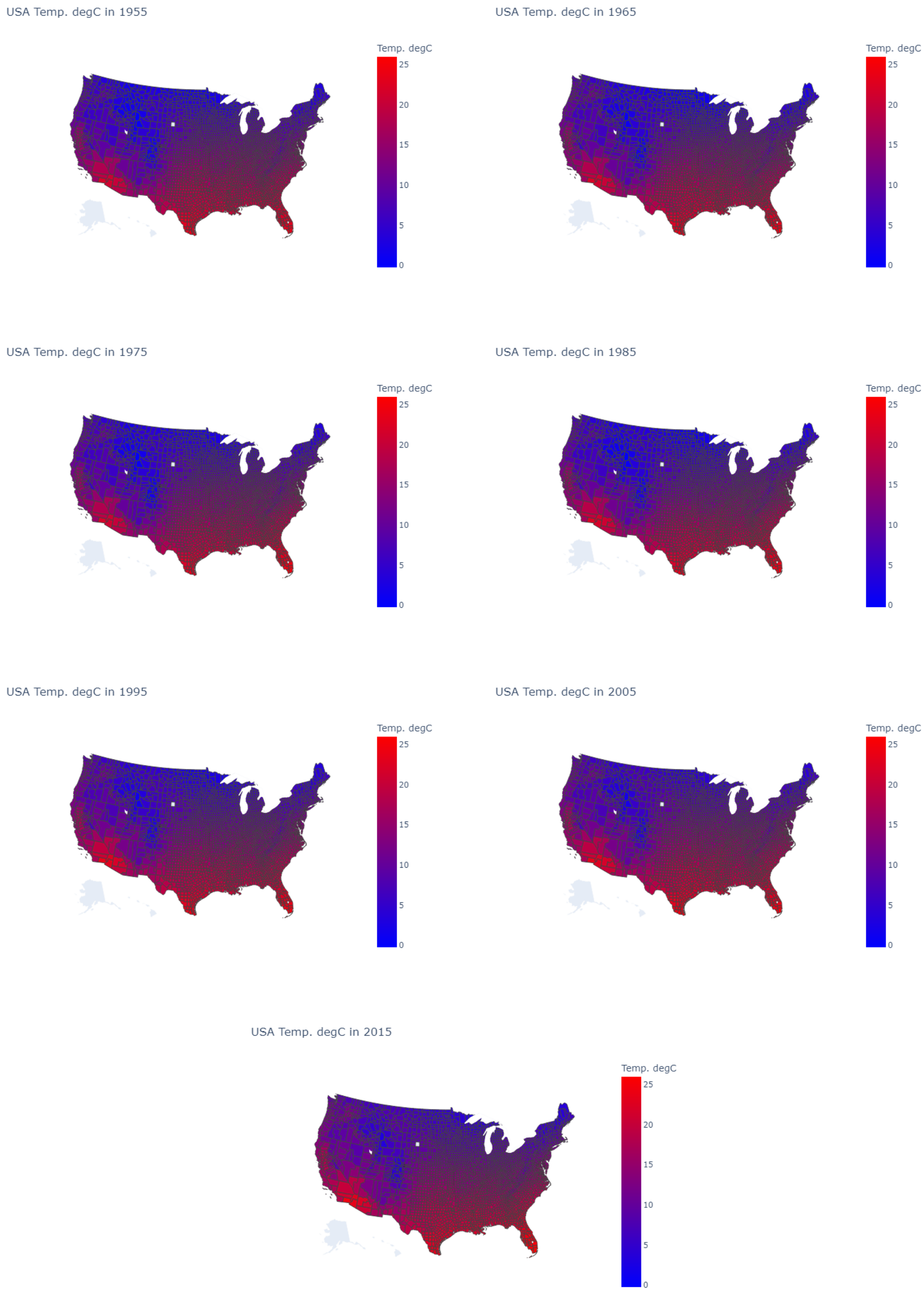


Figure 4: US choropleth yearly temperature maps in 10-year intervals.

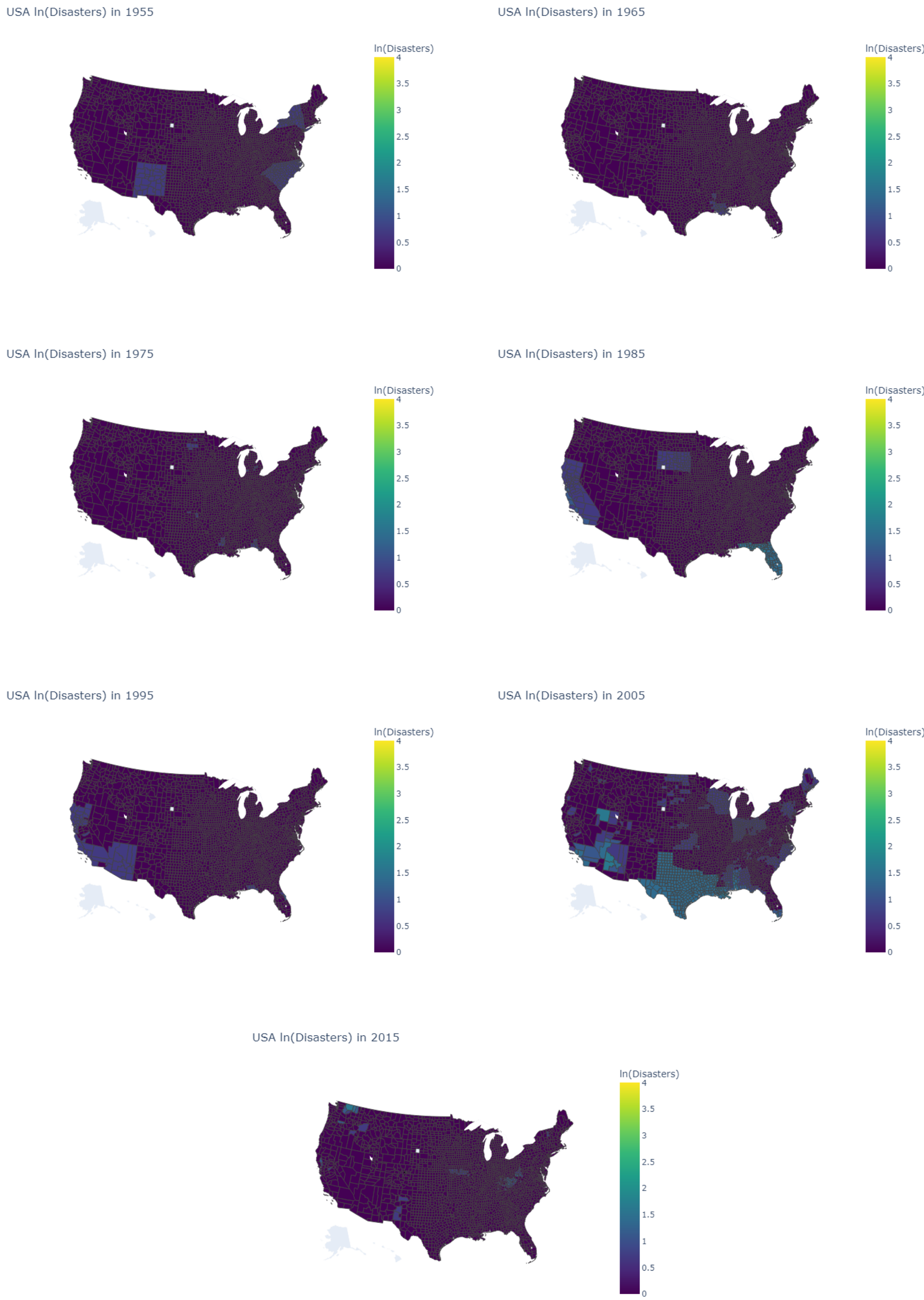


Figure 5: US choropleth yearly logarithm disaster maps in 10-year intervals. The logarithm was taken to narrow the range of values for better visualization.

Identification Strategy

The Ideal Experiment

The ideal experiment to investigate the causal effect of rising global temperatures on the number of disasters is to analyze a large number of *identical* locations. Half of the locations would be subjected to steady average-temperature rise over a long span of time (think: decades) while the control locations would be subjected to perfectly consistent temperature cycles throughout the year. The weather in all the locations would be recorded, and the number of extreme weather events, as decided using a consistent metric. At the end of the experiment, the causal effect of rising temperatures on the number of severe weather events would be clear.

As ideal as this experiment would be, it is not possible to perform. Instead, only collected data can be analyzed, with measures taken to control for selection biases and differences between compared groups.

Issues with a Bivariate OLS

A naive approach to this analysis would be to perform a simple bivariate OLS analysis of the form showed in Eq. (1).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t} + \epsilon_{i,t} \tag{1}$$

where $disasters_{i,t}$ is the number of disasters declared in county i in year t and $tempc_i$ is the temperature, in Celsius, in county i in year t.

This would not be accurate in predicting the causal effect of temperature on the number of severe climate events because it does not control for differences in the locations or in time. In other words, without including the necessary controls, there would be an omitted variable bias (OVB), and the conditional independence assumption (CIA) would *not* hold.

Baseline Regression

To control for the differences in location, a baseline analysis was performed with county fixed effects. The regression equation can be seen in Eq. (2).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t} + \alpha_i + \epsilon_{i,t} \tag{2}$$

where $disasters_{i,t}$ is the number of disasters declared for a given county i in year t, $tempc_{i,t}$ is the average temperature, in Celsius, for county i in year t, and α_i are county fixed effects.

For this regression to show a causal effect, the CIA must hold. It is believed that controls for time are important, but these are not present in this regression. Therefore, this regression will not show a causal effect, but will still be important as a baseline for later regressions.

Baseline Regression with Linear Time

Building on the baseline regression, time, specifically the year, could be used as a regressor to account for changes over time. The regression equation with a linear time trend can be seen in Eq. (3).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t} + \beta_2 t + \alpha_i + \epsilon_{i,t} \tag{3}$$

where $disasters_{i,t}$ is the number of disasters declared for a given county i in year t, $tempc_{i,t}$ is the average temperature, in Celsius, for county i in year t, t is a regressor for the year, and α_i are county fixed effects.

Since this regression accounts for time, it most likely shows a more accurate relationship between temperature and disasters, but may not be causal for not accounting for time fixed effects; this will be discussed next.

Baseline Regression with Time Fixed Effects

Rather than using time as a regressor, it could be used as a fixed effect. By doing this, country-wide year-specific changes could be accounted for. This includes things like federal pollution policy changes ("Pollution Prevention Law and Policies", n.d.; "Regulations for Greenhouse Gas Emissions from Passenger Cars and Trucks", n.d.), which implement various policies, such as limitations on the pollution from known greenhouse gasses like carbon dioxide (Delworth et al., 1999), methane (Smith et al., 2013; Fuglestvedt et al., 1996), and nitrous oxide (Kroeze, 1994).

The time fixed effects are easily included in the baseline regression in the same way as the county fixed effects were included. The regression equation can be seen in Eq. (4).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t} + \alpha_i + \tau_t + \epsilon_{i,t} \tag{4}$$

where $disasters_{i,t}$ is the number of disasters declared for a given county i in year t, $tempc_{i,t}$ is the average temperature, in Celsius, for county i in year t, α_i are county fixed effects, and τ_t are time fixed effects.

Since this regression also accounts for time, it most likely shows a more accurate relationship between temperature and disasters than the baseline, but may not be causal for not accounting for the linear time effects mentioned previously.

Baseline Regression with Both Linear Time and Time Fixed Effects

To account for both linear time effects and time fixed effects, a model that combines both can be run. The regression equation can be seen in Eq. (4).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t} + \beta_2 t + \alpha_i + \tau_t + \epsilon_{i,t} \tag{5}$$

where $disasters_{i,t}$ is the number of disasters declared for a given county i in year t, $tempc_{i,t}$ is the average temperature, in Celsius, for county i in year t, t is a regressor for the year, α_i are county fixed effects, and τ_t are time fixed effects.

With controls for county fixed effects, time fixed effects, and linear time trends, it is believed that all the necessary controls are included to satisfy the CIA. Therefore, the results from this regression would show a causal effect of temperature on severe weather disasters, holding fixed the effects of location (county) and time (year).

Lagged Effects

An additional model was run with lagged effects. The rational behind this model was that temperature rises may have delayed effects on climate, meaning it may take time for temperature rises to cause changes in severe weather events. Introducing a lag would also help combat reverse causality, where, say, an increase in wild fires causes a higher temperature. Due to time constraints, this model was only run with time fixed effects and not both. The regression equation for the lagged effect model can be seen in Eq. (6).

$$disasters_{i,t} = \beta_0 + \beta_1 tempc_{i,t-10} + \alpha_i + \tau_t + \epsilon_{i,t} \tag{6}$$

where $disasters_{i,t}$ is the number of disasters declared for a given county i in year t, $tempc_{i,t-10}$ is the average temperature, in Celsius, for county i in year $t-10$, α_i are county-fixed effects, and τ_t are time-fixed effects.

Like the previous regression in *Baseline Regression with Time Fixed Effects*, this regression includes county and time fixed effects. Assuming lagged effects are important, this regression would show a causal relationship between temperature and severe weather events in the US.

Averaged Data

Because of issues with running the regression on the large dataset, an additional regression model was generated with aggregated data. Investigated was data aggregated into 9- or 10-year groups, averaging the data within that "decade"[1]. The regression equation for the averaged data model can be seen in Eq. (7).

$$disasters_{i,d_k} = \beta_0 + \beta_1 tempc_{i,d_k} + \alpha_i + \tau_d + \epsilon_{i,d_k} \tag{7}$$

where $disasters_{i,d_k}$ is the number of disasters declared for a given county i in decade k, $tempc_{i,d_k}$ is the average temperature, in Celsius, for county i in decade k, α_i are county-fixed effects, and τ_d are time-fixed effects for the decades.

Having removed a lot of the variation in the data through aggregation, this model would not be causal, but would still show accurate trends (i.e. if temperature increases or decreases the number of severe weather disasters per year) because it includes the necessary controls through county and time fixed effects.

Temperature Rise

In order to establish that there is, in fact, a trend of rising temperatures, a temperature versus time regression model was setup. As with previous models in this report, county-fixed effects were used to control for location differences. The regression equation for the temperature versus time model can be seen in Eq. (8).

$$tempc_{i,t} = \beta_0 + \beta_1 year_t + \alpha_i + \epsilon_{i,t} \tag{8}$$

[1] Because the years investigated started in 1953 and ended in 2019, the 67 year range was split into four groups of 10 years and three groups of 9 years. The groups were 1953-1962, 1963-1972, 1973-1982, 1983-1992, 1993-2001, 2002-2010, 2011-2019.

where $tempc_{i,t}$ is the average temperature, in Celsius, for a given county i in year t, $year_t$ is the year, and α_i are county-fixed effects.

For this regression to show the relationship between time and temperature, the CIA must hold. It is believed that county fixed effects are the only good controls for this regression, so there does not exist an OVB. Therefore, the CIA holds, and this regression would show how the average yearly temperature changes over time.

It should be noted that this regression does not imply that time causes temperature rises, rather, through other means, such as pollution, over time causes temperature rise. But, including something like pollution rates would be a bad control because only through pollution would time effect temperature.

Results

Temperature Regression Results

The temperature versus time regression was run in Python using the `OLS` regressor from *statsmodel* (Seabold & Perktold, 2010). The regression was run using robust standard errors (referred to as `HC1` within *statsmodels*), producing heteroskedastically robust standard error values. The model used county-fixed effects through the use of dummy variables for each of the 3,105 counties in the dataset.

The regression results, shown in Table 3, were all statistically significantly different than 0 to a 1% significance level. The results can interpreted to mean that for every 1-year increase, the temperature increases, on average, by $0.02°C$, within each county. This is exactly on par with Lindsey and Dahlman, 2021, showing that there is a trend of the temperature rising (global warming) in the contiguous US.

Table 3: Temperature vs. time regression results with time-invariant fixed effects.

	tempc
const	-30.33***
	(0.20)
year	0.02***
	(0.00)
R-squared	0.98
R-squared Adj.	0.98

Heteroskedastically robust standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

If a $0.2°C$ temperature rise over 1 decade sounds small, it should be noted that the trends in the early 20th century were $0.1°C$ per decade (Lindsey & Dahlman, 2021). This means that the temperature rise is accelerating, and could increase even more if climate change is not curbed.

Baseline Regression

The main baseline regression was run similarly to the temperature vs time regression using an `OLS` regressor in Python. Because this dataset was so large, a computer with a large amount of RAM was needed to run the regeression. The regression was run using robust standard errors, producing heteroskedastically robust standard error calculations. The regression model accounted for county effects through the use of dummies for the 3,105 counties.

The regression results, shown as regression (1) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can interpreted to mean that for every $1°C$ increase in temperature, the number of disaster declarations increases, on average, by 0.18, within each county. Although this may not seem like a large increase, the result are per-county, meaning if the entire US increases in temperature by $1°C$, there will be around 559 (from 3105×0.18) more weather-related disaster declarations per year in the US.

But, this regression is not causal because it does not control for time.

Baseline Regression with Linear Time

Adding to the main baseline, the time regressor was included. The regression was run using robust standard errors, producing heteroskedastically robust standard error calculations. The regression model accounted for county effects through the use of dummies for the 3,105 counties.

The regression results, shown as regression (2) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can interpreted to mean that for every $1°C$ increase in temperature, the number of disaster declarations increases, on average, by 0.12, within each county.

But, this regression is not causal because it does not properly control for time.

Baseline Regression with Time Fixed Effects

Again, the regression was run in Python, but this time with both county and time fixed effects. This was done through the use of 3,105 county dummies and 67 year dummies. The regression was run using robust standard errors, producing heteroskedastically robust standard error calculations.

The regression results, shown as regression (3) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can interpreted to mean that for every $1°C$ increase in temperature, the number of disaster declarations increases, on average, by 0.23, within each county. In other words, if the entire US were to, on average, increase by $1°C$, there would be around 714 (from 3,105 $\times$ 0.23) more severe weather disasters in that year alone.

Baseline Regression with Both Linear Time and Time Fixed Effects

The baseline regression with both linear time trends and time fixed effects was run in Python. The methods for each were outlined above and are not repeated. The regression was run using robust standard errors, producing heteroskedastically robust standard error calculations.

The regression results, shown as regression (4) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can interpreted to mean that for every $1°C$ increase in temperature, the number of disaster declarations increases, on average, by 0.23, within each county holding fixed the effects of time. In other words, if the entire US were to, on average, increase by $1°C$, there would be around 714 (from 3,105 $\times$ 0.23) more severe weather disasters in that year alone.

Interestingly, the coefficient on temperature from this regression was the same as the baseline regression with only time fixed effects, but the constant term changed.

Because this regression includes the necessary controls, this shows the causal effect of temperature on the number of severe weather disasters in the contiguous US.

Lagged Regression Results

The lagged regression model was run with a 10-year temperature lag. This was done by shifting back the temperature data by 10 years and then dropping data points that no longer had temperature data after the shift. The regression was run using robust standard errors (referred to as `HC1` within *statsmodels*), producing heteroskedastically robust standard error calculations. The regression model accounted for county- and time-fixed effects through the use of dummies for the 3,105 counties and 67 years (1953-2019).

The 10-year lag regression results, shown as regression (5) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can be interpreted to mean that for every $1°C$ increase in temperature in year t, the number of disaster declarations in year $t + 10$ increases, on average, by 0.18, within each county and holding fixed the effects of time. In other words, if the temperature increases by $1°C$ in 2021, there will be around 0.18 more severe weather disasters in 2031, within each county and holding fixed the effects of time, or 559 (from 3,105 $\times$ 0.18) more disasters in total.

Averaged Regression Results

The average regression utilized a modified version of the combined dataset which averaged the temperature and disaster columns in ten year intervals. The regression was run with robust standard errors.

The regression results, shown as regression (6) in Table 4, were all statistically significantly different than 0 to a 1% significance. The regression results can interpreted to mean that for average $1°C$ increase within a decade, the number of disaster declarations increases per year by 0.68, within each county and holding fixed the effects of time. In other words, if the average temperature change between 2011 and 2020 was $1°C$, there will have been around around 1800 (from 3105 $\times$ 0.68) more county disaster declarations per year for that 10-year period (or about 18,000 more disasters over that decade).

Note that this analysis method was initially proposed because it aggregates the data, requiring less RAM to run the regression. A consequence of aggregating the data is that it removes some of the variations, making it less accurate. This regression should, therefore, only be used as another proof that temperature and the number of severe disasters are positively correlated.

Regression Summary Table

Table 4: All the regression results summarized together.

	Baseline (1)	Linear Time (2)	Time FE (3)	Both (4)	Lagged (5)	Averaged (6)
const	-2.92***	-13.51***	-4.23***	-0.58***	-3.19***	-11.84***
	(0.13)	(0.29)	(0.31)	(0.02)	(0.20)	(0.96)
tempc	0.18***	0.12***	0.23***	0.23***		0.68***
	(0.01)	(0.01)	(0.02)	(0.02)		(0.05)
tempc(t-10)					0.18***	
					(0.01)	
R-squared	0.04	0.04	0.13	0.13	0.14	0.26
R-squared Adj.	0.02	0.02	0.12	0.12	0.12	0.14
county FE	YES	YES	YES	YES	YES	YES
time FE	NO	NO	YES	YES	YES	YES
linear time	NO	YES	NO	YES	NO	NO
Observations	208034	208034	208034	208034	176984	21735

Heteroskedastically robust standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

Conclusion

The goal of this report was to study the effect of increasing global temperature due to climate change on the number of FEMA-declared, weather-related disasters. The contiguous US county panel temperature and FEMA datasets were analyzed with a temperature versus time regression and six temperature versus disaster regressions, which included a baseline regression with only county fixed effects, a baseline regression with county fixed effects and a linear time regressor, a baseline regression with county and time fixed effects, a baseline with county and time fixed effects and a linear time regessor, a lagged regression with county fixed effects, and an averaged data regression with county fixed effects.

The temperature versus time regression established that there exists a trend of rising temperature during the period studied (1953 - 2019), and confirmed the results of Lindsey and Dahlman, 2021[2] with a statistically significant coefficient[3] of $0.02°C$ per year. This

[2]Note that the results of Lindsey and Dahlman, 2021 found an increase in temperature of $0.2°C$ *per decade*, which aligns with the result of $0.02°C$ per year when it is converted to per decade units.

[3]This means the corresponding p-value was small enough at the 1% significant level to reject the null

means that, on average, an increase in one year results in an increase in $0.02°C$ across the contiguous US within each county.

As discussed in this report, the baseline regression result was not a causal effect of changing temperature on disasters as the regression does not control for time. The two ways to control for time were (1) to add it as an additional linear regressor or (2) add it through time fixed effects. A final baseline regression included both time controls, and produced statistically significant coefficients and was interpreted to mean that for every $1°C$ temperature rise, the number of disasters would, on average, increase by 0.23 disasters, within each county and holding fixed the effects of time.

The lagged effects regression resulted in a statistically significant (at the 1% significance level) temperature coefficient of 0.18 with standard error of 0.01. This was interpreted to mean that a $1°C$ increase in temperature would result in 0.18 more disasters 10 years later, within each county and holding fixed the effects of time. The value is similar to the time fixed effects regression result, but the validity of the regression is unclear.

If a lag is not accurate, then the baseline regression with both linear time trends and time fixed effects would be the most accurate estimate of the causal effect of rising temperature on FEMA-disasters declared in the contiguous US. If the lag is an accurate representation of how temperature effects disasters, then the lagged regression would be the most accurate estimate of the causal effect of rising temperature on FEMA-declared disasters in the contiguous US. Because both regressions produced very similar results, using the lower prediction from the lagged regression would indicate that a $1°C$ increase in temperature would result in around 559[4] more severe weather disasters in the US.

As confirmed by this report, the average US temperature is rising at a rate of $0.02°C$ per year. As a result of this, there will be an increased number of severe weather disasters declared across the contiguous US in the coming years. This will result in enormous loss of property and life. Policies must be enacted to address the accelerating global temperature rise as a result of pollution and harmful human activities or else the US will suffer the ramifications of the temperature rise poses.

Future work should include accounting for the political bias inherent to the FEMA dataset to get a better causal effect of temperature on declared disasters. Additionally, the deaths and property damage resulting from the included disasters should be studied to quantify the effect of increased disasters on the US population.

hypothesis that the coefficient is equal to zero. In other words, the result of $0.02°C/year$ is statistically significantly different from zero.

[4] 3105 counties * 0.18 disasters per county = 559 disasters declared

Appendix

Summary Data

Table 5: Summary statistics of the combined dataset for each year in 5-year steps.

year	tempc_mean	tempc_std	disasters_mean	disasters_std
1953	12.96	4.43	0.16	0.36
1958	11.6	4.32	0.13	0.33
1963	11.93	4.45	0.47	0.58
1968	11.63	4.31	0.02	0.16
1973	12.45	4.35	0.37	0.79
1978	11.38	4.72	0.1	0.31
1983	11.94	4.18	0.08	0.3
1988	12.2	4.23	0.16	0.41
1993	11.73	4.77	0.55	1.07
1998	13.63	4.43	3.43	9.14
2003	12.42	4.45	0.41	0.65
2008	12.17	4.76	0.66	0.91
2013	12.17	4.64	0.38	0.72
2018	12.84	4.74	0.3	0.6

Python Code

Baseline Regression

```python
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

Baseline regression with time trends for the project

Includes county FE

Regression:
    disasters_{i,t} = b0 + b1 tempc_{i,t} + a_i + e_{i,t}
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/CombinedData.csv",
                   usecols=("fips", "tempc", "disasters"))
print("Done.")
```

```python

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1, 1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["fips"], drop_first=True)
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open('../Results/baseline_regression2_summary.txt', 'w') as f:
    f.write("Baseline Regression 2")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Baseline Regression with Linear Time

```python
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

Baseline regression with time trends for the project

Includes county FE and linear time trends

Regression:
    disasters_{i,t} = b0 + b1 tempc_{i,t} + b2 year + a_i + e_{i,t}
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/CombinedData.csv",
                   usecols=("fips", "year", "tempc", "disasters"))
```

```python
print("Done.")

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1, 1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["fips"], drop_first=True)
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open('../Results/baseline_regression_with_time_trends_summary.txt',
          'w') as f:
    f.write("Baseline Regression with Time Trends")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Baseline Regression with Time Fixed Effects

```python
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

Baseline regression with time FE for the project

Regression:
    disasters_{i,t} = b0 + b1 tempc_{i,t} + a_i + T_t + e_{i,t}
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/CombinedData.csv",
                   usecols=("fips", "year", "tempc", "disasters"))
```

```python
print("Done.")

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1, 1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["year", "fips"], drop_first=True)
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open('../Results/baseline_regression_with_timeFE_summary.txt', 'w') as f:
    f.write("Baseline Regression with Time FE")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Baseline Regression with Both Linear Time and Time Fixed Effects

```python
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

Baseline regression with time trends and time fixed effects for the project

Includes county FE, linear time trends, and time fixed effects

Regression:
    disasters_{i,t} = b0 + b1 tempc_{i,t} + b2 t + a_i + T_t + e_{i,t}
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/CombinedData.csv",
```

```
                   usecols=("fips", "year", "tempc", "disasters"))
print("Done.")

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1, 1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["year", "fips"], drop_first=True)
X["year"] = data.year
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open('../Results/baseline_regression_with_both_summary.txt',
          'w') as f:
    f.write("Baseline Regression with Both")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Lagged Effects Regression

```
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

Lagged effects
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/CombinedData.csv",
                    usecols=("fips", "year", "tempc", "disasters"))
print("Done.")
```

```python

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1,1)
print("Done.")


lag = -10
print(f"Lagging {lag} years...", end="")
data.tempc = data.tempc.shift(lag)
print("Done.")


print("Dropping data...", end="")
max_year = data.year.max()
drop_indx = data.year > (max_year + lag)
print(f"Dropped {sum(drop_indx)} rows...", end="")
data = data[~drop_indx]
y = y[~drop_indx]
print("Done.")


print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["year", "fips"], drop_first=True)
print("Done.")


print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")


print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open(f'../Results/lagged_effects_{lag}years_summary.txt', 'w') as f:
    f.write(f"Lagged Effect {lag} Years Regression")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Averaged Data Regression

```python
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
```

```
Fall 2021
SS340 Cause and Effect
Final Project

Averaged-data regression for the project
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/AveragedData.csv",
                    usecols=("fips", "years", "tempc", "disasters"))
print("Done.")

print("Creating endogenous variable...", end="")
y = data.pop("disasters")
y = y.to_numpy().reshape(-1,1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["years", "fips"], drop_first=True)
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "tempc"],
                            drop_omitted=True)

with open('../Results/averaged_data_regression.txt', 'w') as f:
    f.write("Averaged-Data Regression")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

Temperature vs Time Regression

```
# -*- coding: utf-8 -*-
"""
Jared Jacobowitz
Fall 2021
SS340 Cause and Effect
Final Project

```

```python
Temperature vs. Time regression to show that the temperature is increasing
"""
import pandas as pd
import statsmodels.api as sm
from statsmodels.iolib.summary2 import summary_col

print("Loading data...", end="")
data = pd.read_csv("../Datasets/TemperatureData.csv",
                    usecols=("fips", "year", "tempc"))
data = data[(data.year >= 1963) & (data.year <= 2019)]
print("Done.")

print("Creating endogenous variable...", end="")
y = data.pop("tempc")
y = y.to_numpy().reshape(-1,1)
print("Done.")

print("Creating Dummies...", end="")
X = pd.get_dummies(data, columns=["fips"], drop_first=True)
print("Done.")

print("Running Regression...", end="")
X_sm = sm.add_constant(X)
model = sm.OLS(y, X_sm).fit(cov_type="HC1")
print("Done.")

print("Saving Summary...", end="")
stata_summary = summary_col(model,
                            stars=True,
                            float_format="%0.2f",
                            regressor_order=["const", "year"],
                            drop_omitted=True)

with open('../Results/tempvstime_regression_summary.txt', 'w') as f:
    f.write("Temperature vs Time Regression")
    f.write('\n')
    f.write(stata_summary.as_text())
print("Done.")
```

References

NASA. (2021). Climate Change Evidence: How Do We Know? https://climate.nasa.gov/evidence/

Lindsey, R., & Dahlman, L. (2021). Climate Change: Global Temperature. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

FEMA. (2021a). Disaster Information. https://www.fema.gov/disaster

FEMA. (2021b). Disaster Declarations Summaries - V2. https://www.fema.gov/openfema-data-page/disaster-declarations-summaries-v2

GitHub, & The Washington Post. (2021). The Washington Post's Analysis of NOAA Climate Change Data for the Contiguous United States. https://github.com/washingtonpost/data-2C-beyond-the-limit-usa

USDA. (n.d.). County FIPS Codes. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697

Kiprop, J. (2018). What Are the Primary Causes of Dam Failure? https://www.worldatlas.com/articles/what-are-the-primary-causes-of-dam-failure.html

Center for Climate and Energy Solutions. (2021). Drought and Climate Change. https://www.c2es.org/content/drought-and-climate-change/

Tschumi, E., & Zscheischler, J. (2019). Countrywide Climate Features During Recorded Climate-Related Disasters. *Climatic Change*, *158*(3-4), 593–609. https://doi.org/10.1007/s10584-019-02556-w

Michener, W. K., Blood, E. R., Bildstein, K. L., Brinson, M. M., & Gardner, L. R. (1997). Climate Change, Hurricanes and Tropical Storms, and Rising Sea Level in Coastal Wetlands. *Ecological Applications*, *7*(3), 770–801. https://doi.org/10.2307/2269434

Wasko, C., Sharma, A., & Pui, A. (2021). Linking Temperature to Catastrophe Damages from Hydrologic and Meteorological Extremes. *Journal of Hydrology*, *602*, 126731. https://doi.org/10.1016/j.jhydrol.2021.126731

Flannigan, M. D., Amiro, B. D., Logan, K. A., Stocks, B. J., & Wotton, B. M. (2006). Forest Fires and Climate Change in the 21st Century. *Mitigation and Adaptation Strategies for Global Change*, *11*(4), 847–859. https://doi.org/10.1007/s11027-005-9020-7

Banholzer, S., Kossin, J., & Donner, S. (2014). The Impact of Climate Change on Natural Disasters. *Reducing Disaster: Early Warning Systems For Climate Change*, 21–49. https://doi.org/10.1007/978-94-017-8598-3_2

Environmental Protection Agency. (n.d.). Climate Change and Harmful Algal Blooms. https://www.epa.gov/nutrientpollution/climate-change-and-harmful-algal-blooms

Brooks, H. (2013). Severe Thunderstorms and Climate Change [6th European Conference on Severe Storms 2011. Palma de Mallorca, Spain]. *Atmospheric Research*, *123*, 129–138. https://doi.org/https://doi.org/10.1016/j.atmosres.2012.04.002

Irland, L. C. (2000). Ice Storms and Forest Impacts. *Science of The Total Environment*, *262*(3), 231–242. https://doi.org/https://doi.org/10.1016/S0048-9697(00)00525-8

Karl, T. R., & Trenberth, K. E. (2003). Modern Global Climate Change. *Science*, *302*(5651), 1719–1723. https://doi.org/10.1126/science.1090228

Fassnacht, S. R., Cherry, M. L., Venable, N. B. H., & Saavedra, F. (2016). Snow and Albedo Climate Change Impacts Across the United States Northern Great Plains. *The Cryosphere*, *10*(1), 329–339. https://doi.org/10.5194/tc-10-329-2016

NASA, & Buis, A. (2019). Can Climate Affect Earthquakes, Or Are the Connections Shaky? – Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/news/2926/can-climate-affect-earthquakes-or-are-the-connections-shaky/

Pollution prevention law and policies. (n.d.). https://www.epa.gov/p2/pollution-prevention-law-and-policies

Regulations for Greenhouse Gas Emissions from Passenger Cars and Trucks. (n.d.). https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-greenhouse-gas-emissions-passenger-cars-and

Delworth, T. L., Mahlman, J. D., & Knutson, T. R. (1999). Changes in Heat Index Associated with CO2-Induced Global Warming. *Climatic Change*, *43*(2), 369–386. https://doi.org/10.1023/a:1005463917086

Smith, K. R., Desai, M. A., Rogers, J. V., & Houghton, R. A. (2013). Joint CO2 and CH4 Accountability for Global Warming. *Proceedings of the National Academy of Sciences*, *110*(31), E2865–E2874. https://doi.org/10.1073/pnas.1308004110

Fuglestvedt, J., Isaksen, I., & Wang, W. (1996). Estimates of indirect global warming potentials for CH4, CO and NOX. *Climate Change*, *602*, 405–437. https://doi.org/10.1007/BF00139300

Kroeze, C. (1994). Nitrous Oxide and Global Warming. *Science of The Total Environment*, *143*(2), 193–209. https://doi.org/https://doi.org/10.1016/0048-9697(94)90457-X

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.