

# Are increased incidence of extreme storm events associated with a change in homeowner insurance costs?

Ryan Altman\*

Ehsan Habibpour<sup>†</sup>

Natalie Short<sup>‡</sup>

2025-01-11

## Abstract

Using data from the United States Census Bureau and the National Oceanic and Atmospheric Administration, we examine whether the incidence of extreme storm events is associated with a change in reported homeowner insurance costs.

[1] 4560

## 1 Introduction

In recent years, the homeowners insurance market has been disrupted by more extreme weather events - increasing both in storm intensity and incidence - which has made the provision of this insurance unprofitable in many states, resulting in rate hikes or the cessation of coverage altogether (Flitter (2024a)). Major insurers State Farm, Farmers, and Allstate have all pulled out of providing homeowners insurance in California. State Farm, previously the largest homeowners insurer in the state, cited “historic increases in construction costs outpacing inflation, rapidly growing catastrophe exposure, and a challenging reinsurance market” when they announced this change in 2023 (Farm 2023). In this study, we search for evidence of the second reason by examining whether severe storm event data is correlated with homeowners reported insurance costs.

## 2 Literature Review

Since extreme weather events are expected to increase in intensity and incidence as climate change worsens, this issue will likely only get worse (Climate Change (IPCC) 2021). However, it is difficult to obtain data on homeowners insurance costs, as state insurance regulators have neglected to collect this data from insurance companies, so it is not possible to see exactly how much these premiums have changed over time, besides relying on anecdotal reports from homeowners (Flitter 2024b). (Kousky2019role?) in a review of disaster insurance markets, points out that while natural disaster losses have been increasing worldwide, there is surprisingly little empirical research on how these events affect insurance costs. This limitation in research is particularly notable when

---

\*American University Undergraduate Student

<sup>†</sup>American University Graduate Student

<sup>‡</sup>American University Graduate Student

studying the relationship between extreme weather events and homeowner insurance costs, making it difficult to understand how increases in storm frequency might influence insurance markets over time.

In order to integrate citations into the References section below, we add entries into our file `main.bib`. This is a plain-text file that we edit in RStudio (or BibDesk, or similar). We store `main.bib` in the same folder as our paper’s `.qmd` and `.pdf` files. Its entries are formatted so that they can be knit to `.pdf`; see <https://j.mp/2UzTXEZ> for example entries for articles, books, and miscellaneous. We can get these entries automatically from Google Scholar by turning on BibTeX in the Google Scholar Settings - Bibliography Manager. Perhaps we use a tool like free, open-source BibDesk to help us manage the `.bib` file.

### 3 Data

We choose two separate datasets to examine this problem. To estimate homeowner insurance costs, we use the American Community Survey, which is part of the US Census and is administered five years. Our unit of analysis is county-year, to estimate the variation in monthly homeowner costs across geography and time.

Our county-level census data comes from the American Community Survey (ACS) 5-year estimates, accessed through the Census Bureau’s API. We collect data for three distinct, yearlong periods: 2013, 2018, and 2023. For each county, we extract several housing-related variables including the median monthly owner costs, median house value, median household income, and average household size. The county identification is based on FIPS (Federal Information Processing Standards) codes, which consist of a two-digit state code and a up-to-three-digit county code.

To ensure comparability across different time periods, all monetary values are adjusted for inflation using price indices (with 2023 as the base year, applying indices of 0.765 for 2013 and 0.824 for 2018). Monthly owner costs are annualized by multiplying by 12. The final dataset is created by combining data from all three time periods and organizing it by state, county, and year.

Our dataset includes several key variables: total number of housing units, units with and without mortgages, average household size, median household income, median house value, and median monthly owner costs (separated into categories for properties with and without mortgages). We also collect data on owner costs as a percentage of yearly income, both overall and separated by mortgage status. County names are standardized by removing suffixes such as “County,” “Census Area,” “Municipio,” “Parish,” “Borough,” “Planning Region,” and “City” to ensure consistency across different naming jurisdictions.

It is important to note that while the 2023 ACS included a separate variable for homeowner insurance, this is not available in some years’ surveys. In surveys prior to 2023, homeowner fire, hazard, and flood insurance costs are incorporated into a broader measure called “Selected Monthly Owner Costs,” which also encompassed real estate taxes, mortgages, utilities, and fuels. To account for these varying components, we include house value in our model as a proxy for real estate taxes and mortgage costs, while household size is included to capture variation in utility and fuel costs.

Our weather event data is obtained from the National Oceanic and Atmospheric Administration (NOAA) through Google Cloud’s BigQuery platform. We access the public dataset ‘bigquery-public-data.noaa\_historic\_severe\_storms.storms’ using Python’s BigQuery client library. To establish the connection, we create a service account and associated credentials file that allows secure access to the Google Cloud API.

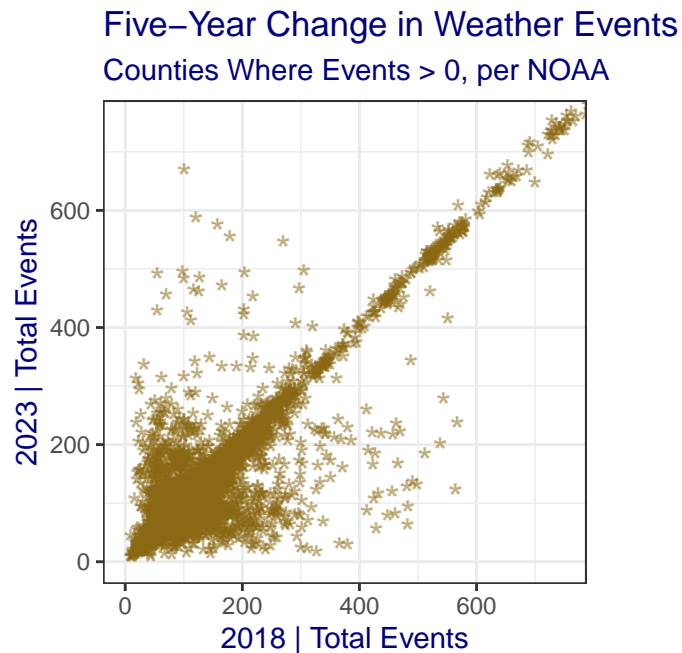
The query extracts yearly event data at the county level, including state and county FIPS codes for precise geographic identification. For each county, we collect the number of occurrences of different

types of severe weather events. The initial format of the NOAA data presents individual reports of specific weather events by county and year (for example, one row would read as “King County, Washington had 4 reports of debris flow in 2013”). To make this data more amenable to analysis, we transform it using a `pivot_wider()` function to create a format where each county-year combination represents a single row. We then aggregate all event reports for each county-year into a total event count, which serves as our primary measure of severe weather activity in our analysis.

The data is organized by year, state, and county, with corresponding FIPS codes to ensure accurate matching with our census data. The query groups events by these geographic identifiers and event type, providing a comprehensive count of different severe weather occurrences for each location and time period. This structured approach allows us to create a consistent dataset that can be merged with our housing cost data for subsequent analysis.

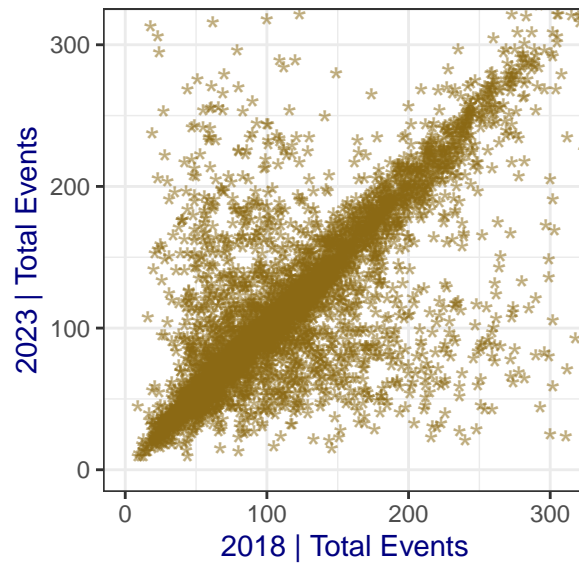
#### Exploratory Data Analysis

Are storms correlated by county?



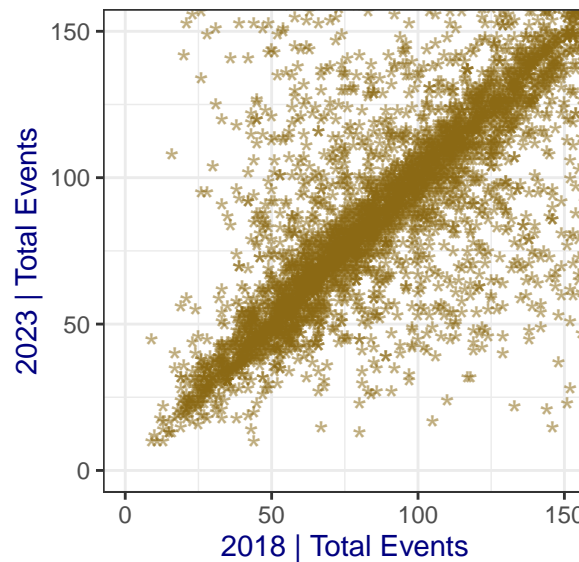
Let's zoom in on the majority of counties:

### Five-Year Change in Weather Events Counties Where Events > 0, per NOAA



...and zoom in even closer:

### Five-Year Change in Weather Events Counties Where Events > 0, per NOAA



A linear model for this relationship:

Call:  
`lm(formula = noaa_2023_bothyearcounties$total_events ~ noaa_2018_bothyearcounties$total_events)`

Residuals:

Min	1Q	Median	3Q	Max
-389.17	-11.26	-2.20	12.00	567.32

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.13071	1.14049	12.39	<2e-16
noaa_2018_bothyearcounties\$total_events	0.88659	0.00618	143.45	<2e-16

(Intercept) \*\*\*  
noaa\_2018\_bothyearcounties\$total\_events \*\*\*  
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54.36 on 5341 degrees of freedom  
Multiple R-squared: 0.7939, Adjusted R-squared: 0.7939  
F-statistic: 2.058e+04 on 1 and 5341 DF, p-value: < 2.2e-16

2018 events and 2023 events are highly correlated by county, which makes sense because weather is generally stable within a 5-year period. Overall, though, there is a very slight positive relationship between year-change and total events – demonstrating climate change and the greater prevalence of extreme weather events.

Delineating by event type:

Floods:

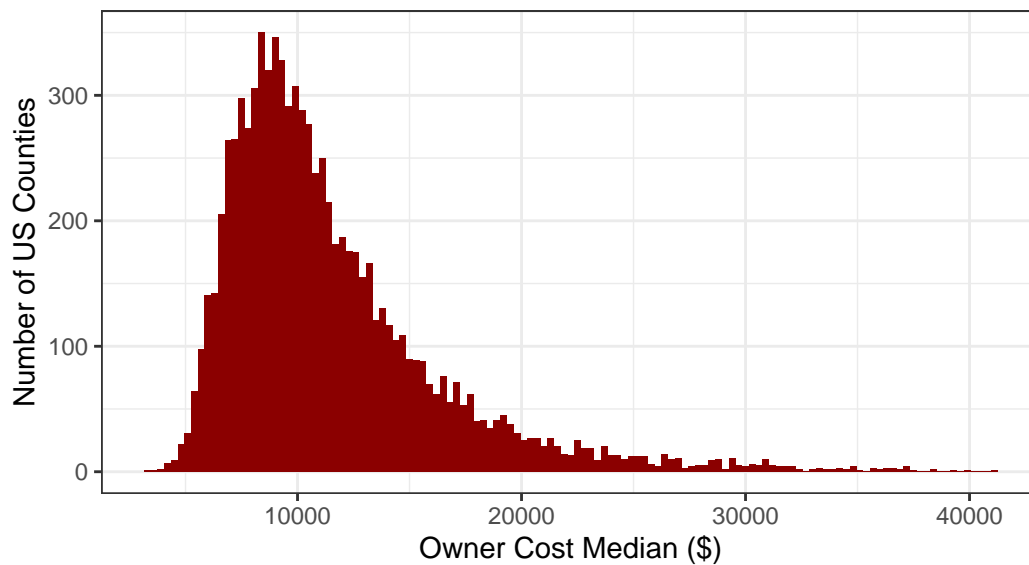
event_type	instances
Length:4	Min. : 13.0
Class :character	1st Qu.: 479.5
Mode :character	Median :21250.0
	Mean :22862.8
	3rd Qu.:43633.2
	Max. :48938.0

```
# A tibble: 4 x 2
  event_type instances
  <chr>      <dbl>
1 flash flood 48938
2 flood      41865
3 coastal flood 635
4 lakeshore flood 13
```

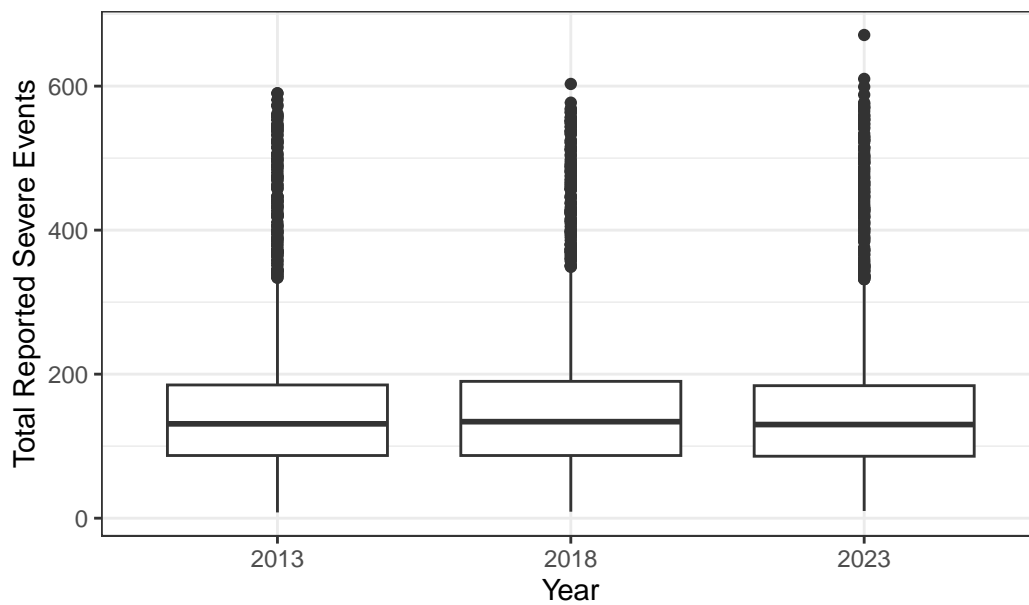
[1] 1059

More EDA:

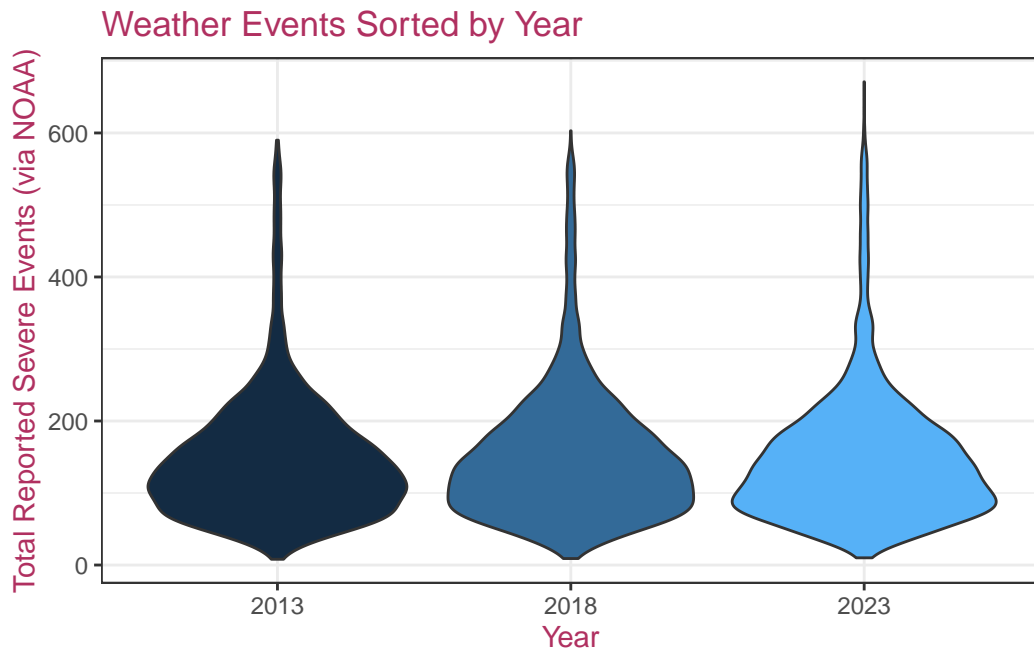
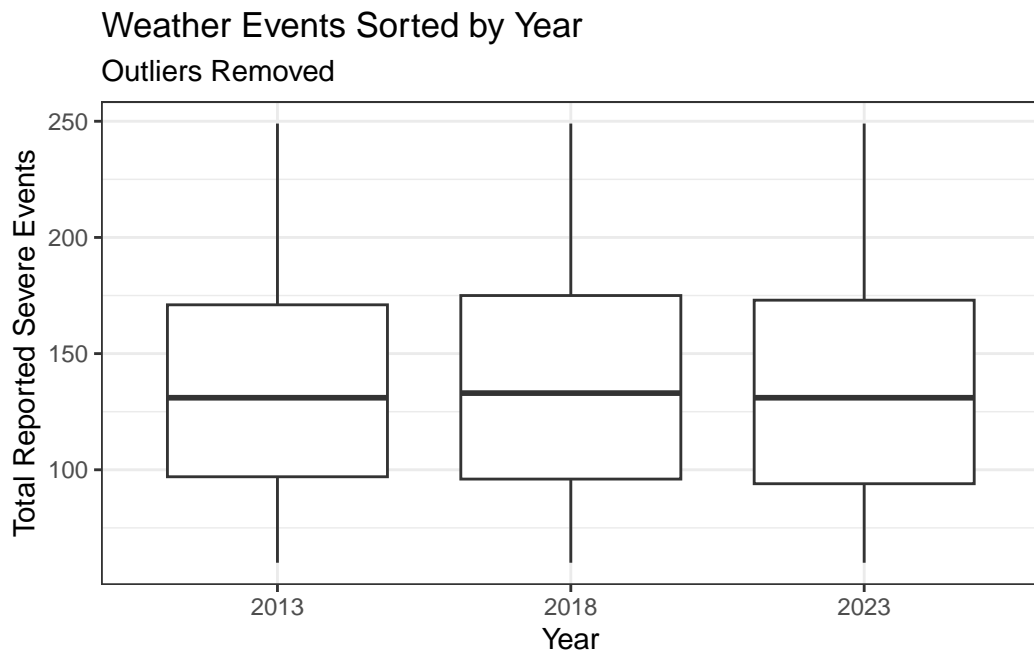
Median of Estimated Monthly Homeowner Costs  
(with insurance)



Weather Events Sorted by Year



Warning: Removed 1674 rows containing non-finite outside the scale range  
(`stat\_boxplot()`).



## 4 Methods

We merge NOAA data and ACS data using a Full Join, based on a combination of county names and Federal Information Processing Standard (FIPS) geographic codes. This allows for counties with the same name but different geographies (for example, there are 25 Jefferson counties in the United States) to not merge any data.

Once we have this main, joined dataset, we are able to run statistical analyses. Wanting to check 2013 weather events' possible effects on 2018's homeowner costs in Clark County, Kentucky – as one of many examples – we also make a shifted dataset. This set includes a new variable that applied a positive 5-year shift on NOAA data, allowing for easy modeling based on – again, as one example – 2018 weather reporting on 2023 costs by county.

The unshifted linear model uses the unshifted weather data to predict housing costs in the same year: 2013 costs by 2013 weather events in each county, 2018 costs by 2018 weather events in each county, and 2023 costs by 2023 weather events in each county. The model begins with an intercept of \$11,777.40 and predicts a decrease of \$1.96 in annual housing costs for each extreme weather event in the county. This model is statistically significant, with a p-value of 0.0007; the linear model has an adjusted R-squared that is effectively 0.

The shifted linear model uses the shifted dataset to predict housing costs 5 years later: 2018 costs by 2013 weather events in each county and 2023 costs by 2018 weather events in each county. The model begins with an intercept of \$11,351.34 and predicts a decrease of \$1.31 in annual housing costs for each extreme weather event in the county five years previous. This model is not statistically significant, with a p-value of 0.0536; the model also has an adjusted R-squared that is approaching 0.

We also estimate a fixed effect model to explore how homeowner costs relate to natural disaster events within each state over time. Our regression equation is structured as:

$$Y_{ti} = \beta X_{ti} + \text{fixed effects}$$

In this equation,  $Y_{ti}$  represents the logarithm of median owner cost in county  $i$  at time  $t$ . The explanatory variables  $X_{ti}$  include our main variable of interest - the logarithm of total number of events - along with the following control variables: the logarithm of median house value and the average household size at the county level. The fixed effects component captures county-specific characteristics that remain constant over time. By using fixed effects at the county level, we control for time-invariant factors specific to each county that might influence homeowner costs.

This model specification allows us to examine how changes in natural disaster events within a county relate to changes in homeowner costs, while controlling for other relevant factors and accounting for county-specific characteristics. The use of logarithmic transformations helps us interpret the relationships in terms of percentage changes and helps address potential non-linearities in the relationships between variables.

## 5 Linear Model

```
model_1 <- lm(owner_cost_median ~ total_events, data = main_set)
summary(model_1)
```

Call:

```
lm(formula = owner_cost_median ~ total_events, data = main_set)
```

Residuals:

Min	1Q	Median	3Q	Max
-8290	-3228	-1256	1804	29762



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11777.3958	101.6632	115.847	< 2e-16 ***
total_events	-1.9624	0.5813	-3.376	0.00074 ***

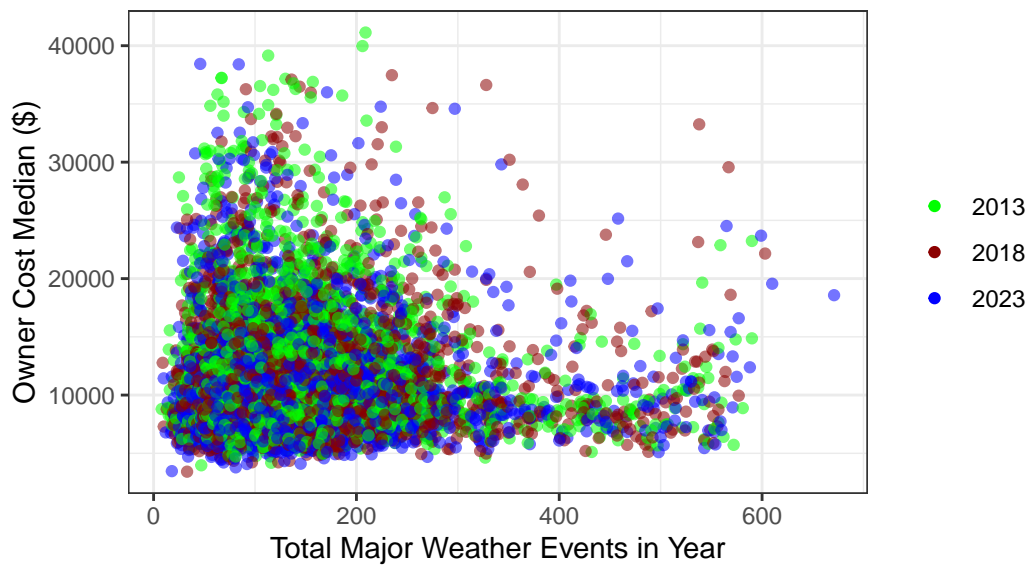
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4876 on 8469 degrees of freedom  
Multiple R-squared: 0.001344, Adjusted R-squared: 0.001226  
F-statistic: 11.4 on 1 and 8469 DF, p-value: 0.0007397

A graphical representation of the above:

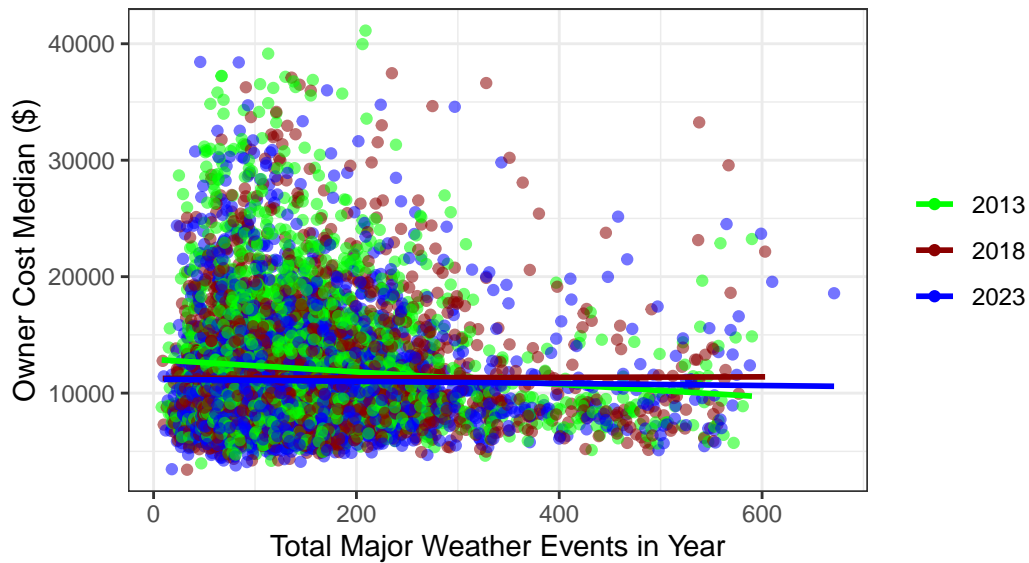
### Major Weather Events' Effect on Costs By U.S. County, By Year



How about we fit a line of best fit on that?

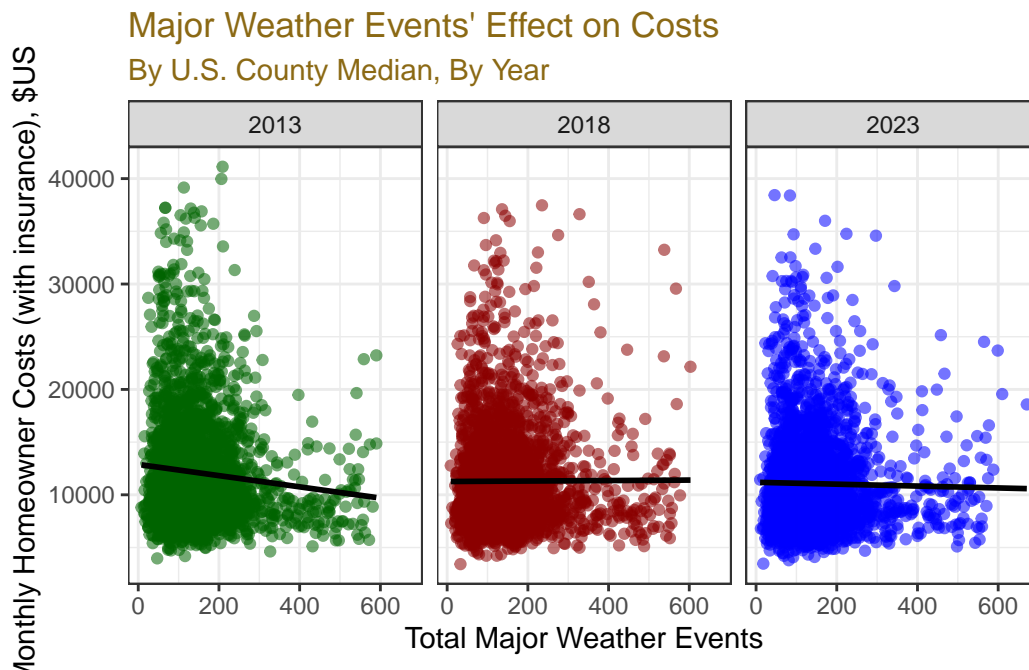
```
`geom_smooth()` using formula = 'y ~ x'
```

## Major Weather Events' Effect on Costs By U.S. County, By Year



Split the above plot into years:

```
`geom_smooth()` using formula = 'y ~ x'
```



Shifting NOAA years to better plot effects YoY:

```
Call:
lm(formula = owner_cost_median ~ total_events, data = main_set_shifted)
```

Residuals:

Min	1Q	Median	3Q	Max
-7867	-3082	-1135	1768	27238

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11351.3419	119.0855	95.32	<2e-16 ***
total_events	-1.3122	0.6799	-1.93	0.0537 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

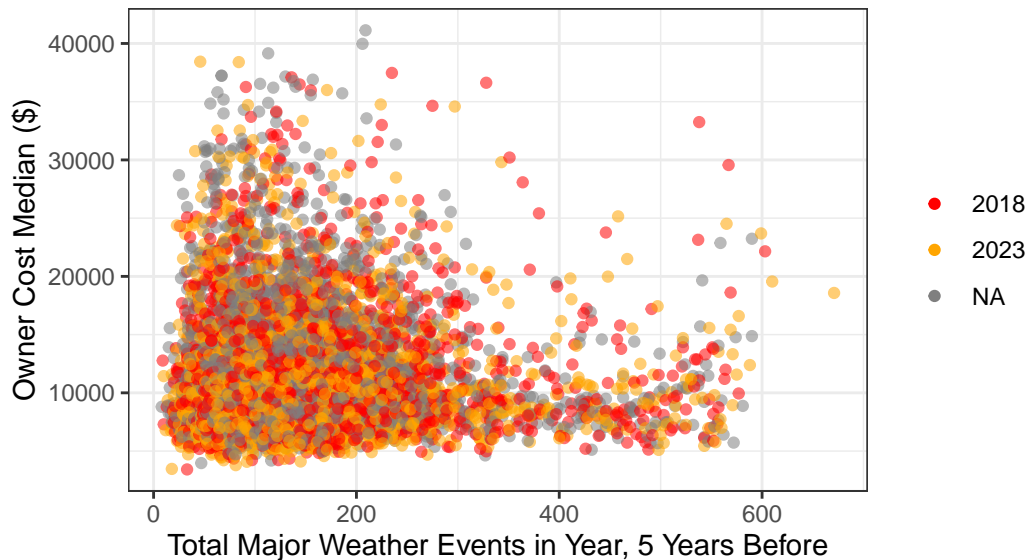
Residual standard error: 4626 on 5634 degrees of freedom

Multiple R-squared: 0.0006607, Adjusted R-squared: 0.0004833

F-statistic: 3.725 on 1 and 5634 DF, p-value: 0.05366

## Major Weather Events' Effect on Costs

By U.S. County, By Year



## 6 Results

Our analysis using fixed effects models reveals some patterns in how natural disasters relate to homeowner costs. The `fe_model2` shows a slight positive relationship between the logarithm of total events and homeowner costs when looking at changes within counties, though this relationship is not statistically significant. Interestingly, `fe_model2` also shows that the log of median house value has a negative correlation with owner costs, which contradicts what prior research would typically expect.

When we examine `fe_model4`, which uses the lag of the logarithm of total events to predict the current logarithm of median owner cost, we find a negative relationship. This unexpected finding suggests that either our model specification is not capturing the true relationship or that there are underlying factors we have not considered. It is important to note that using the lagged variable in `fe_model4` results in losing approximately one-third of our data points (2013 housing data and 2024 weather event data), while still maintaining a separate predictor for each county in our fixed effects estimation.

Our LASSO analysis provides additional insights. In this model, the `log_total_events` variable shrinks to zero, indicating it may not be as strong a predictor as initially thought. The LASSO approach estimates most of the fixed effects for the counties, suggesting that county-specific characteristics play a significant role in determining homeowner costs. This comprehensive analysis raises questions about the true nature of the relationship between natural disasters and homeowner costs, and suggests we might need to explore additional factors or alternative model specifications.

Panel

Fixed effects model

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )
<code>log_total_events_lagged</code>	-0.0124127	0.0072225	-1.7186	0.0858
<code>log_house_value_median</code>	0.0126130	0.0114667	1.1000	0.2714

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Lasso model

Presentation

#### Fixed Effects Regression Results

```

=====
                        Dependent variable:
                        -----
                        Log Owner Cost
                        -----
Log Total Events                -0.012
                                (0.007)

Log House Value                 0.013

-----
County Fixed Effects            Yes
Year Fixed Effects              No
Clustered SE                    County
Observations                    5,569
R2                              0.001
Adjusted R2                     -1.055

```

F Statistic	1.564 (df = 2; 2707)
=====	
Note:	*p<0.05; **p<0.01; ***p<0.001

## 7 Discussion

Our analysis concludes that natural disasters and extreme weather events are not major statistical predictors of homeowner costs and/or insurance pricing. This analysis does come with some caveats, such as the intent and quality of NOAA’s reporting and the lack of an insurance-only variable in the Census data – so it is not a thorough analysis of a possible relationship.

It is possible that the relationship between storm events and homeowner costs is being obscured by variation across counties. Counties with relatively milder climates and fewer storms where we perhaps would not expect as much of a change in homeowner costs, are a high proportion of our dataset. Perhaps further work could investigate outlier counties with more extreme weather events to see if there is a relationship to homeowner costs, or the analysis could be done at the state level. It could also be interesting to compare the counties in the storms dataset to counties that experienced no extreme storm events, or to group the counties in the dataset by number of storms to compare counties with less storms against counties with more storms.

Future work could include an investigation of NOAA’s data, delineating housing costs by ZIP code instead of by county, and the incorporation of county-wide economic data to use as control variables in modeling.

Future work could also explore this problem by storm type, explain changes in storm type incidence and perhaps evaluate whether certain storm types are related to any changes in homeowner insurance costs. Isolating types of storms that are more likely to cause structural damage to homes, such as floods, wildfires and hurricanes, could perhaps improve upon our results.

Future work would benefit from more precise and complete data on current and historical homeowner insurance costs; the same benefit would apply to weather event data.

## References

- Christopher Flavelle, Ivan Penn, Jill Cowan. 2023. “Climate Change, Insurance, and Wildfires in California.” 2023. <https://www.nytimes.com/2023/05/31/climate/climate-change-insurance-wildfires-california.html>.
- Christopher Flavelle, Mira Rojanasakul. 2024. “How Climate Change Is Affecting Insurance for Homes.” 2024. <https://www.nytimes.com/interactive/2024/05/13/climate/insurance-homes-climate-change-weather.html>.
- Climate Change (IPCC), Intergovernmental Panel on. 2021. “Chapter 11: The Earth’s Climate System.” 2021. <https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-11/>.
- Darmiento, Laurence. 2024. “California’s Insurance Crisis: What Went Wrong, What’s Being Done to Fix It, and How Homeowners Can Help Themselves.” 2024. <https://www.latimes.com/business/story/2024-03-29/californias-insurance-crisis-what-went-wrong-whats-being-done-to-fix-it-and-how-homeowners-can-help-themselves>.
- Farm, State. 2023. “State Farm General Insurance Company, California New Business Update.” 2023. <https://newsroom.statefarm.com/state-farm-general-insurance-company-california-new-business-update/>.
- Flitter, Emily. 2024a. “Homeowners Insurance Premiums and Coverage: What You Need to Know.” 2024. <https://www.nytimes.com/2024/02/16/business/homeowners-insurance-bills-coverage.html>.
- . 2024b. “NAIC’s Hurdles in Addressing Homeowners Insurance Crisis.” 2024. <https://www.nytimes.com/2024/03/21/business/naic-homeowners-insurers-states-hurdle.html>.