Methods/Data

National Oceanic and Atmospheric Administration

In this study I am operationalizing extreme weather through data obtained from the National Oceanic and Atmospheric Administration (NOAA, recently rebranded the NCEI). The NOAA contains a comprehensive dataset of weather ranging from the 1950s to now. The data is collected by the National Weather Service (NWS) and is compiled with support from local and national NCEI climate centers. They collect a range of datasets including extreme weather events which they define as “the occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce.” Furthermore, the organization also documents events such “as record maximum or minimum temperatures or precipitation that occur in connection with another event” and track weather events perceived as “rare” weather phenomena that generate media attention.

These data are significant due to the organization’s documentation of “rare” weather events that often receive media attention, increasing the likelihood that they are seen by locals within the same state. This type of extreme weather data is a departure from the data used by the past literature that looks at abnormal weather only operationalized through abnormal temperatures (Bergquist, & Warshaw, 2017; Brooks, Oxley, Vedlitz, Zahran, Lindsey, 2014). That data allows for a new angle on the types of weather that affect climate change attitudes. This data is available for download through the NOAA website on the Storm Events Database page. For the purposes of this study, I have collected only certain types of extreme weather events, although up to 48 are available. I have designated the extreme weather events pertinent to this study as those designated by the National Climate Assessment Council (NCA, 2014):

excessive heat, flash floods, flood, heat, heavy rain, tropical storms, and hurricanes (see Figure 1 for composition of events). I run initial analyses on the relationship between national incedences of extreme weather for investigating I also analyze these extreme weather events both as a total number of extreme weather events for each state over time and by each type of weather event for each state over time (see Figure 2 months with the most extreme and least extreme weather). I analyze three states with the highest amount of extreme weather and three states with the lowest amount of extreme weather in order to discover any ideological differences in the interaction between extreme weather and Google search data. I take one conservative state, one liberal state, and one swing state to represent each group of extreme weather (e.g., the liberal state with the highest amount of extreme weather events and the liberal state with the lowest amount of extreme weather events) (see Figures 3 and 4 for the compositions of each of these groups).

I calculate conservativism/liberalism by taking state-level data from the two closest presidential elections: 2008 and 2010. States with the highest combined voting percentage for the conservative candidate were coded as conservative, those with the lowest were coded as liberal, and those within .05 of 1 (i.e., .5 percent on average each year) were coded as swing states. The swing states included in the state-level analyses for this study are Missouri (extreme weather events 2010-2012= 1910) and Florida (n = 466). The conservative states included are Kentucky (n = 1658) and Alaska (n = 48), and the liberal states are Illinois (n = 1603) and Rhode Island (n=39). This data has recently been used by Konisky and Hughes (2015) in investigating survey reports data and extreme weather event data.

Cooperative Congressional Election Study

I also analyze the Cooperative Congressional Election Study data to replicate the model of a study by Konisky and Hughes (2015) which also analyzed the relationship between abnormal weather and political attitudes. I replicate their model without the Google search data and, I then include Google correlate terms into the OLS methodology they implemented to demonstrate how Google search data is able to improve the predictive power of their model. This data is collected in two waves during election years: pre-election and post-election, in non-election years the data is simply collected once a year. I use the years 2010-2012 to recreate the models used by Konisky (2015). The central dependent variable from this survey data is climate change concern, measured by a 5-point likert scale asking about participant's feelings about climate change.

Google Search Data

I implement the usage of both Google Correlate and Google Trends Data for this study. Google Correlate is able to find Google querries that are correlated with other Google querries and can also find queries that are correlated with user-supplied data. In this instance, I load the compiled extreme weather NOAA data into Google Correlate and use the returned correlates to build upon the model of Konisky. As many of the correlates are likely be spurious, I use stepwise regression to help choose the correct correlates to include in the modified model. Any initial trends in Google search terms are also visible here.

I also implement the use of Google Trends for this study. Google Trends produces an index of search activity for a given search term as opposed to absolute volume of search term. The index of a query is calculated by dividing the number of searches for the desired query by the total number of general Google searches. Moreover, this means that the data given by Google Trends for any given search term is relational – a state with an index of 25 for Google search "weather" has a quarter of the searches during that time as the state with an index of 100 (the max score). Together these data have gained popularity in recent years, used by researchers to both improve extant models and contribute to literature in which it may be difficult to measure a certain behavior for fear of self-censoring – as used in this study (Ginsberg, Mohebbi, Patel, R. S., Brammer, Smolinski, & Brilliant, 2008; King, 2011; Stephens-Davidowitz, 2017). I determine the Google Search Terms used in Google Trends with the use of SEMRush data cross referenced with other Google search volume 3rd party sites. I used these sites to designate the most popular terms used to Google search information about climate change being nonexistent and the most popular result is climate change hoax (see Figure 5). I use the query climate change hoax as a covariate responsible for increases in Google Searches for climate change in conservative states during periods of extreme weather.

Models

For this study I implement several modeling techniques. To first replicate the model of Konisky (2015) with Google Correlate terms I use simple OLS modeling. I also implement stepwise regression to choose correlates from Google Correlate to include in this model. For the main findings of the paper I implement Auto-regressive Integrated Moving Average (ARIMA) modeling to first establish a strong model with the extreme weather data, and then apply this model to the Google Trends data. I end up choosing a (1,1,0) ARIMA model – indicating a 1st order auto-regressive model. Finally, I make use of Bayesian Structural Time Series (BSTS) modeling to investigate causal mechanisms in a way that ARIMA is unable to, that is, BSTS allows for modeling of both regression components and time series components.

Initial Results

I first design my model using the NOAA extreme weather data and model it using ARIMA methodology. The time series for this analysis is constricted to 2010-2012 to compare with the model of Konisky (2015) but will be eventually be expanded to years 2004-2017 in further analyses. Initial graphs of the data suggested a seasonal trend in the data (Figure 2) and initial time-series plotting confirms this (Figure 6). Using decomposition I decompose the seasonal effect and "de-seasonalize" the data (Figure 7). Running an Augmented Dickey-Fuller (ADF) test on this data determines whether it is stationary or not and results indicate *p=.082* indicating a failure to reject the null, confirming a non-stationary time-series.

Test Statistic	Lag Order	P-Value	Alternative Hypothesis
-3.359	3	.08184	Stationary

Graphing the auto-correlation function (ACF) plot (Figure 8) illustrates that there several significant auto correlations (past the blue dotted line) but these auto correlations could be driven by the strong correlation of the first two lags, as seen in the partial-ACF plot (Figure 9). I difference the time-series by d=2 and produce a *p=.04* on another ADF test, rejecting the null and confirming a stationary time-series. I finally fit an ARIMA model with (1,1,0) using the auto.arima() function resulting in a model of: $\hat{Y}_{dt}=0.4226Y_{t-1}+\mathbb{E}$ and summary stats:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
9.77779	53.56872	39.13891	.8285617	3.126874	.1345383	-.0192799

This model can be improved upon due to a less than ideal Mean Absolute Percentage of Error, but this can be solved easily by tinkering with different iterations of the ARIMA model. This model can than be applied to Google Trends data to measure fit and test hypotheses.

Appendices

Figure 1

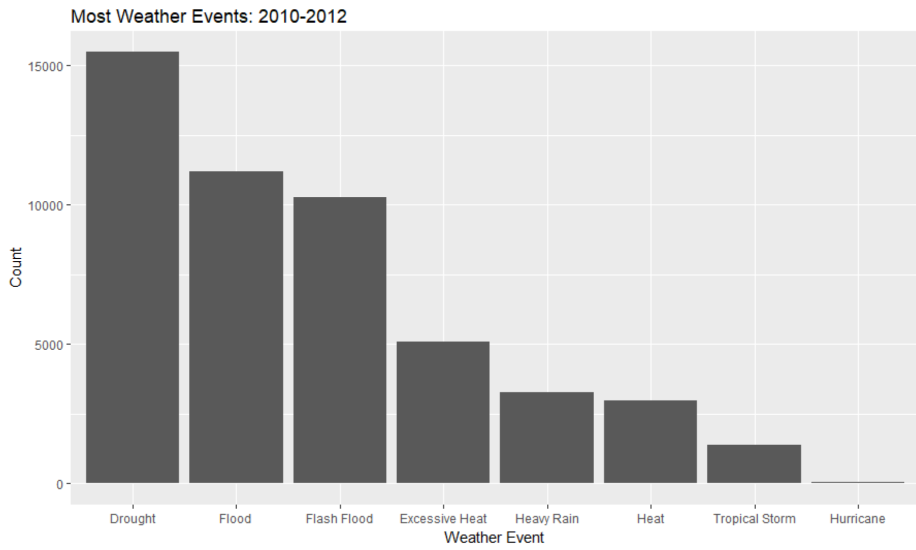

Figure 2

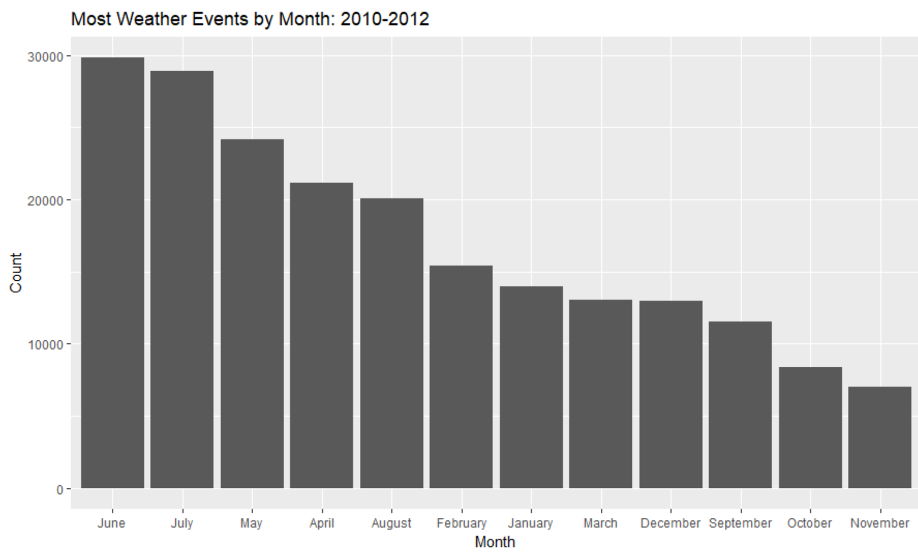

Figure 3

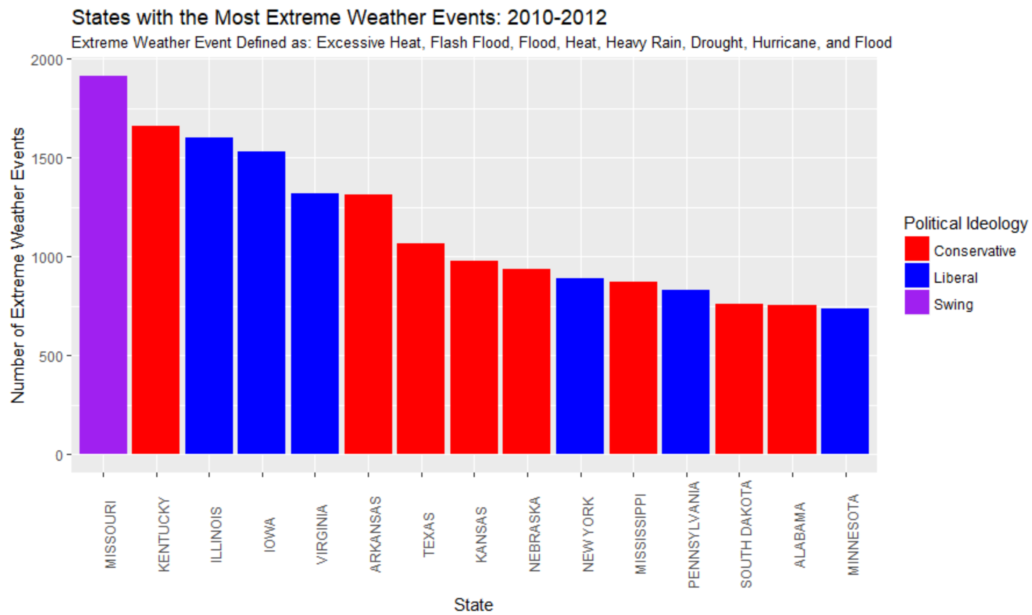

Figure 4

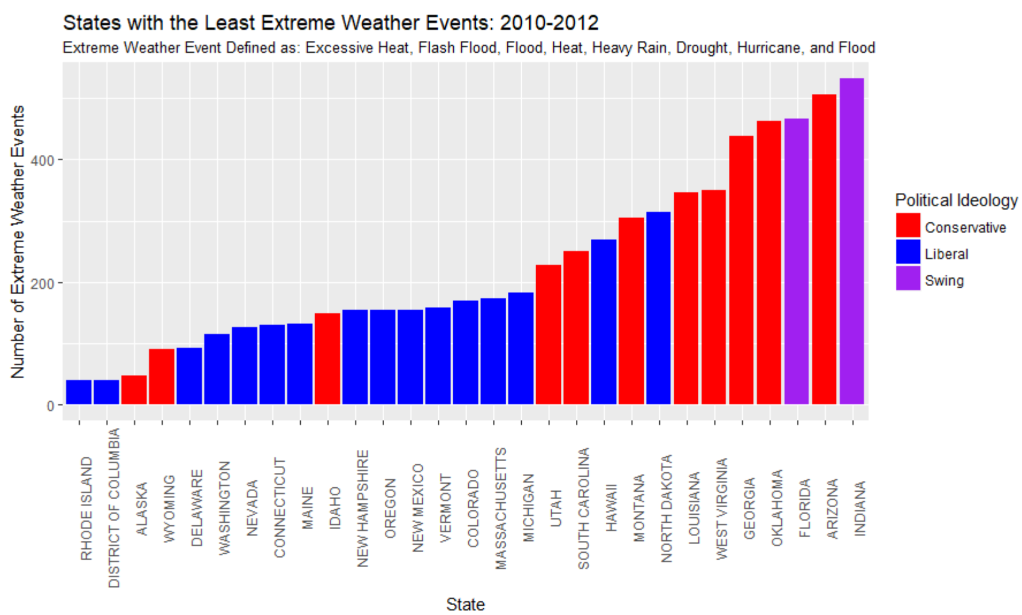

Figure 5

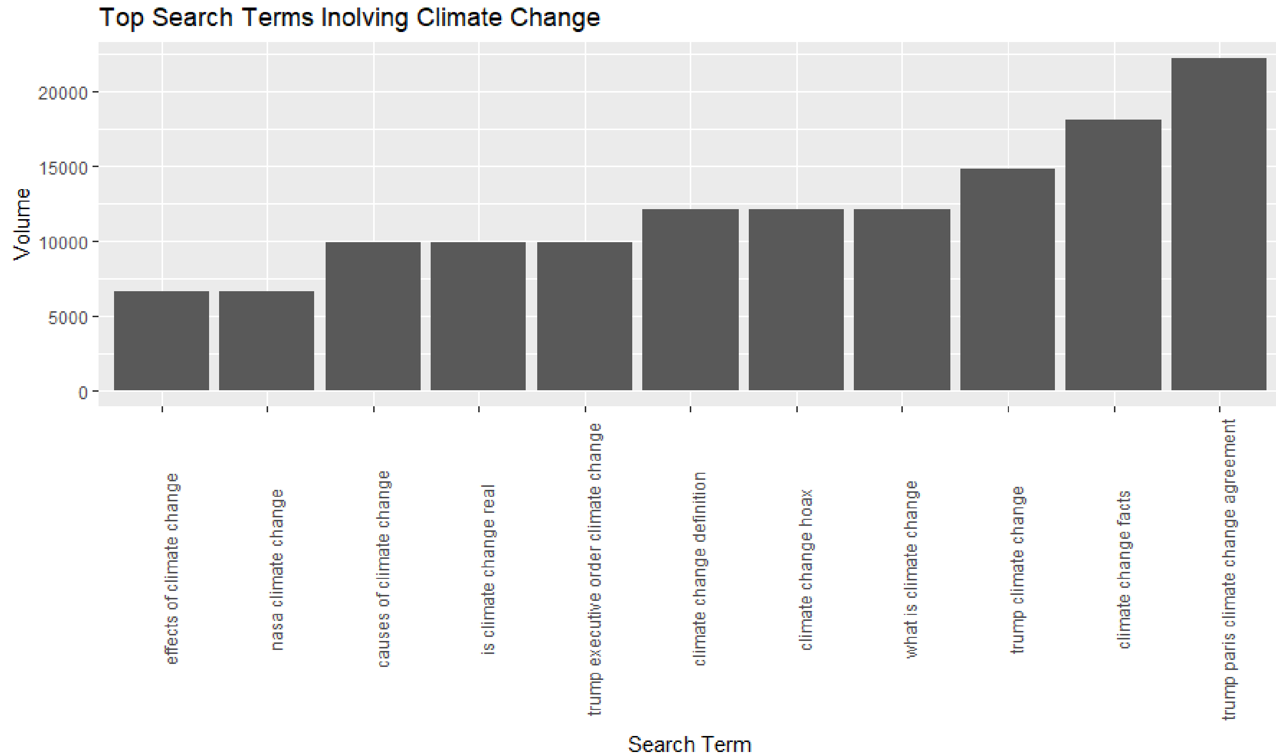

Figure 6

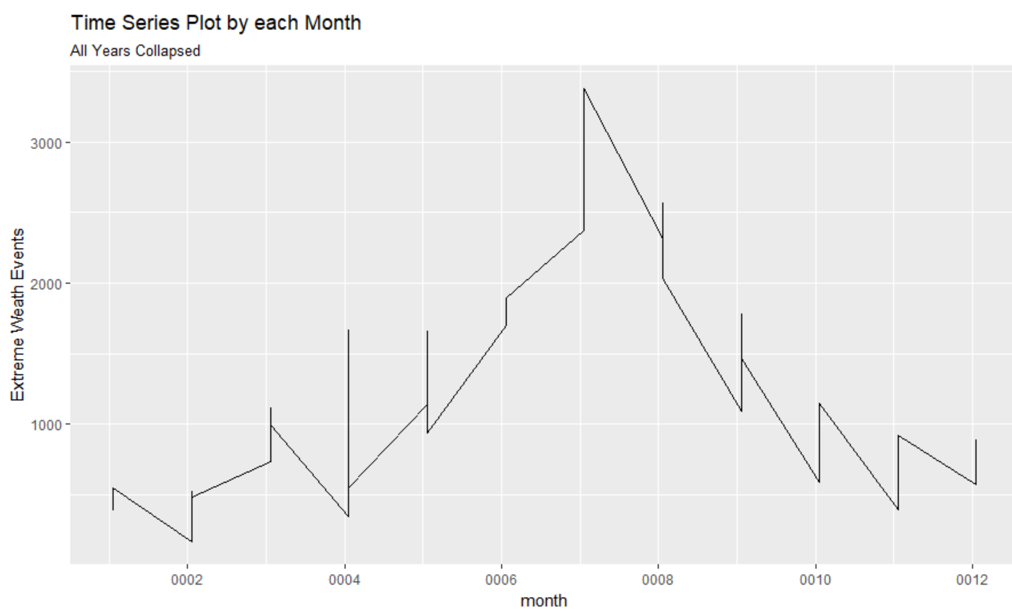

Figure 7

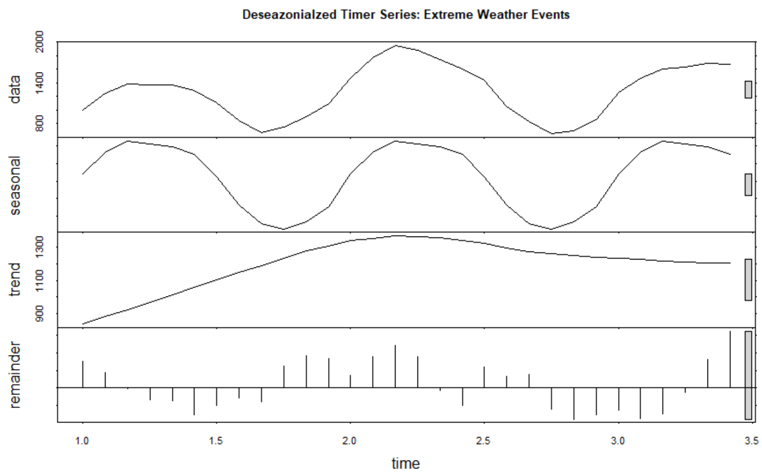

Figure 8

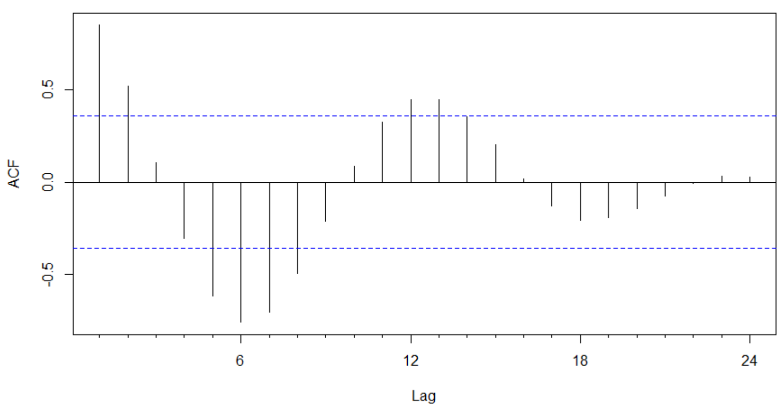

Figure 9

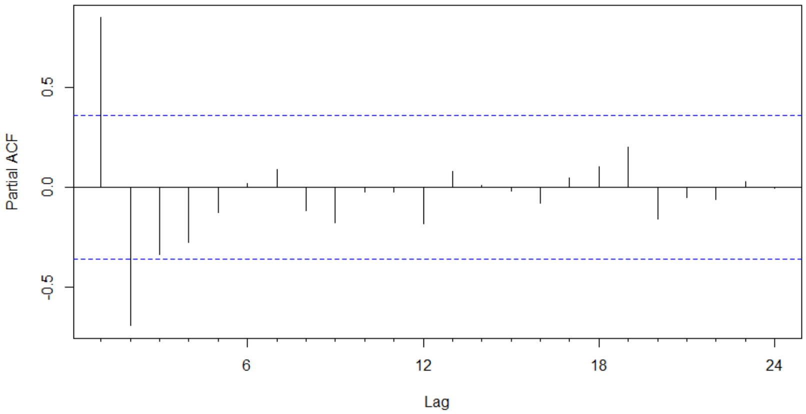

Figure 10