The Effects of Global Warming on Economic Costs in the United States (1980- 2011)

Liam Donovan, Corey Smith, Olivia Sousa

Research Question

How has the **air temperature** of the United States Changed? Is this related to the **severity** of natural disasters?

- Temperature data sourced from the Centers for Disease Control and Prevention (CDC)
- Ranges from 1979 - 2011
- Consists of average low and high temperatures per month, per state (Contiguous United States).
  - From which we extrapolated to find the average temperature (mean of low and high); National Weather Service
- Disaster data sourced from the National Oceanic and Atmospheric Administration (NOAA); National Centers for Environmental Information (NCEI)
  - Ranges from 1980-2011
  - Number of Occurrences (per year), Cost (in billions USD) (also includes confidence intervals of cost)

Data Scraping

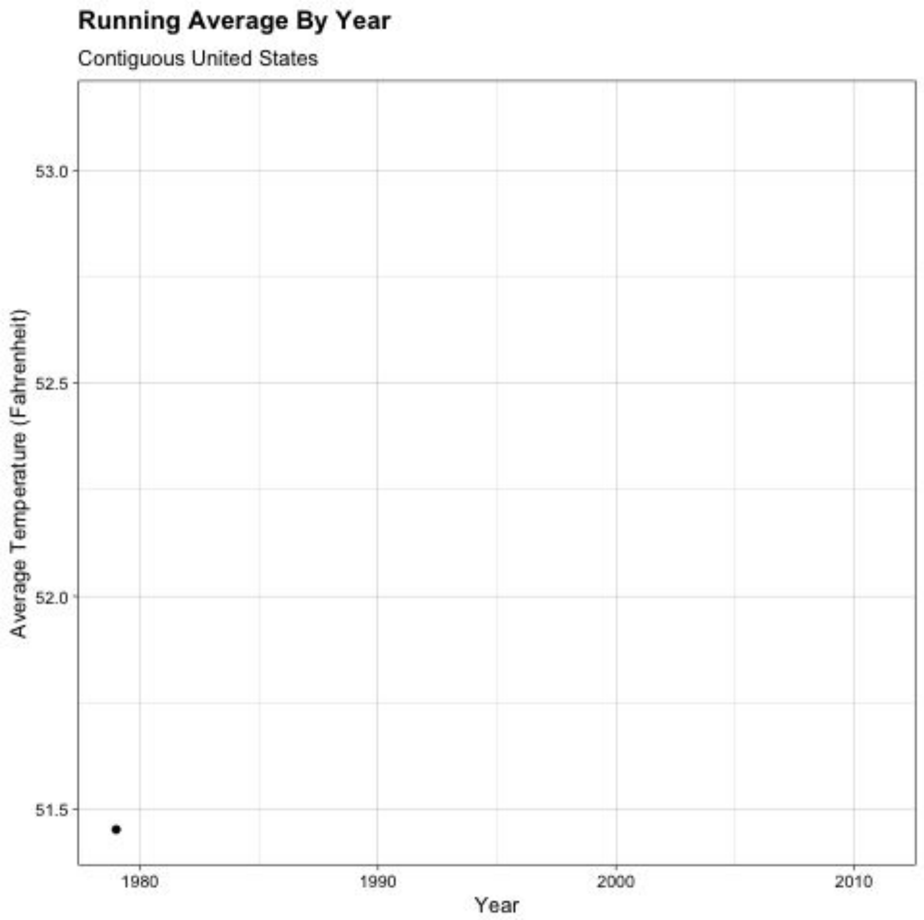
- CDC data given as .txt
  - Can easily be read into a dataframe
- NOAA data given as .xml
  - Must be data scrapped

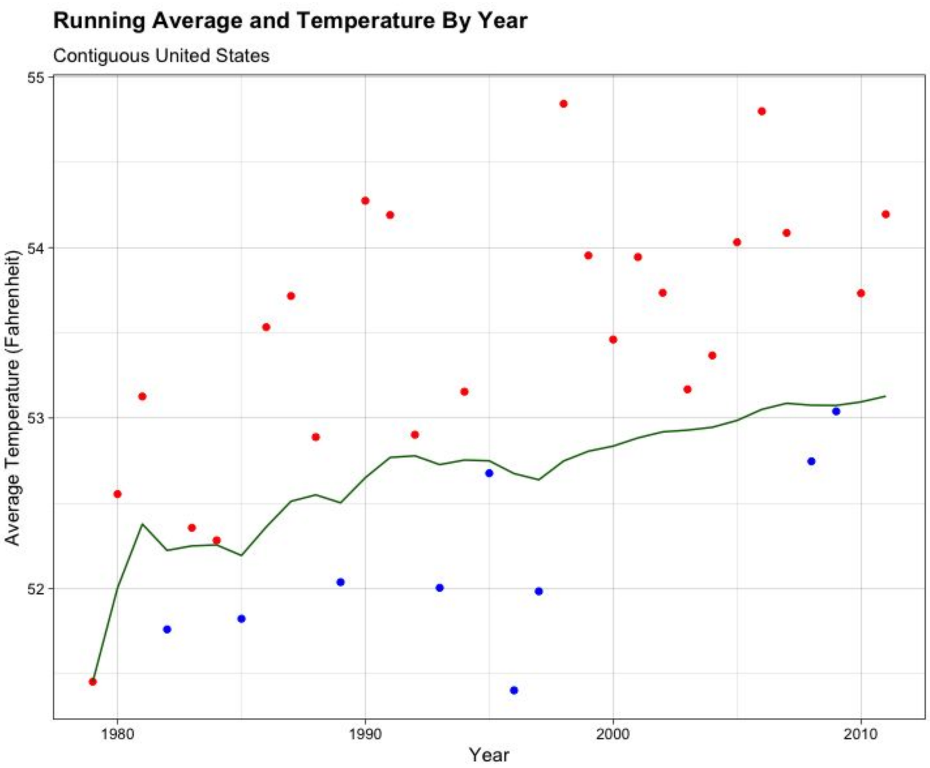
```
 [1] "<?xml version=\"1.0\" encoding=\"UTF-8\"?>"
 [2] "<dataCollection>"
 [3] "  <description>"
 [4] "    <title>United States Billion-Dollar Disasters By Year (CPI-Adjusted)</title>"
 [5] "    <units>Cost values are in billions of dollars</units>"
 [6] "    <additionalInfo>Upper and Lower 75th, 90th, and 95th confidence cost intervals are provided</additionalInfo>
 [7] "  </description>"
 [8] "  <data>"
 [9] "    <year>1980</year>"
[10] "    <drought>"
[11] "      <count>1</count>"
[12] "      <cost>39.7</cost>"
[13] "      <lower75>31.7</lower75>"
[14] "      <upper75>47.5</upper75>"
[15] "      <lower90>29.3</lower90>"
```

	year	disaster	count	cost	lower75	upper75	lower90	upper90	lower95	upper95
1	1980	drought	1	39.7	31.7	47.5	29.3	49.9	28.1	51.1
2	1981	drought	0	0	0	0	0	0	0	0
3	1982	drought	0	0	0	0	0	0	0	0
4	1983	drought	1	9.3	6.6	10.7	6	12	5.4	12.6
5	1984	drought	0	0	0	0	0	0	0	0
6	1985	drought	0	0	0	0	0	0	0	0
7	1986	drought	1	5.1	4.2	5.9	3.9	6.5	3.7	6.7
8	1987	drought	0	0	0	0	0	0	0	0
9	1988	drought	1	53.2	40.4	64.6	37.5	67.6	36.2	69.2
10	1989	drought	1	7.6	6.6	8.9	6.4	9.1	6.1	9.4
11	1990	drought	0	0	0	0	0	0	0	0
12	1991	drought	1	6.9	6.9	8.9	6.6	9.4	6.4	9.6
13	1992	drought	0	0	0	0	0	0	0	0
14	1993	drought	1	2.7	2.8	3.2	2.6	3.5	2.6	3.5
15	1994	drought	0	0	0	0	0	0	0	0
16	1995	drought	1	2	1.8	2.2	1.6	2.6	1.4	2.8
17	1996	drought	1	3.6	3.4	4.6	3.4	5	3.2	5.2
18	1997	drought	0	0	0	0	0	0	0	0
19	1998	drought	1	6.7	5.7	7	5.5	7.4	5.3	7.6
20	1999	drought	1	4.7	4.3	5	4.1	5	4.1	5.2

Overall Air Temperatures Patterns

- Average Temp. per Year
- Temp. has a high variation;
  - Abnormally cold summers, abnormally warm winters, may not be part of a trend
  - Plotting Avg. Temp. has low correlation
- ⇒ Rolling Average (cumulative average)
  - If Avg. is consistently higher than current Avg., will increase





≥ Running Mean

≤ Running Mean

Note: significantly more points above the cumulative average; increasing avg.

Problem:
States have radically different average temperatures. How do we know that this trend is consistent?

Subset by Avg. Temperature

- Compute avg. of each state
- Subset by quantile:
  - Very low, low, high, very high

0%	25%	50%	75%	100%
41.77268	47.73006	52.25905	58.65326	71.88314

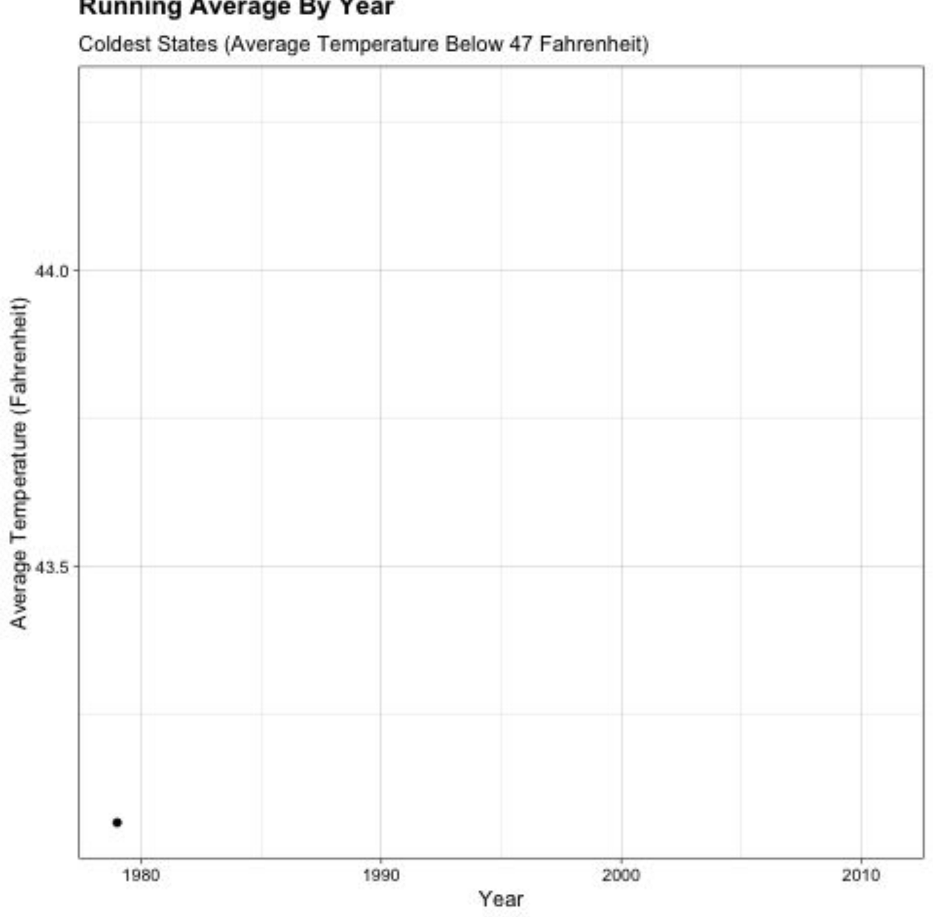
- Interested in extremes (very high/low)
  - Least variation in temp.
  - Consistent change; not random

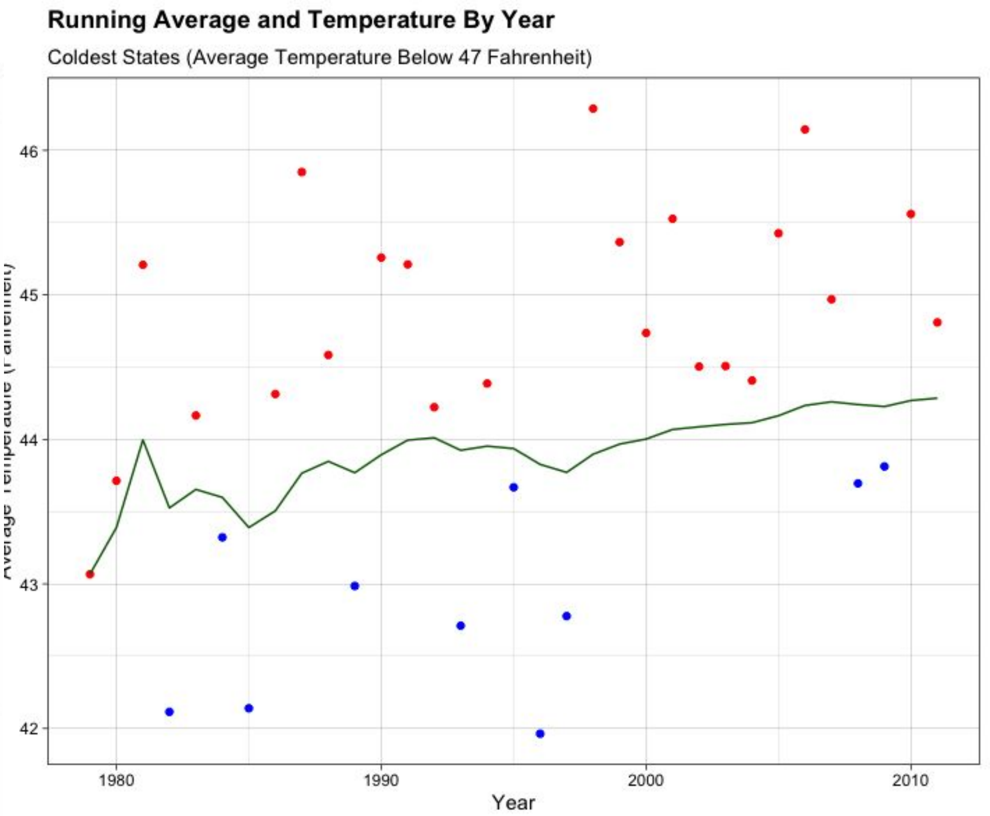
Very High
(quantile 1
Below 47.73)

Colorado
Idaho
Maine
Michigan
Minnesota
Montana
New Hampshire
New York
North Dakota
Vermont
Washington
Wisconsin
Wyoming

Very Low
(quantile 4
Above 58.65)

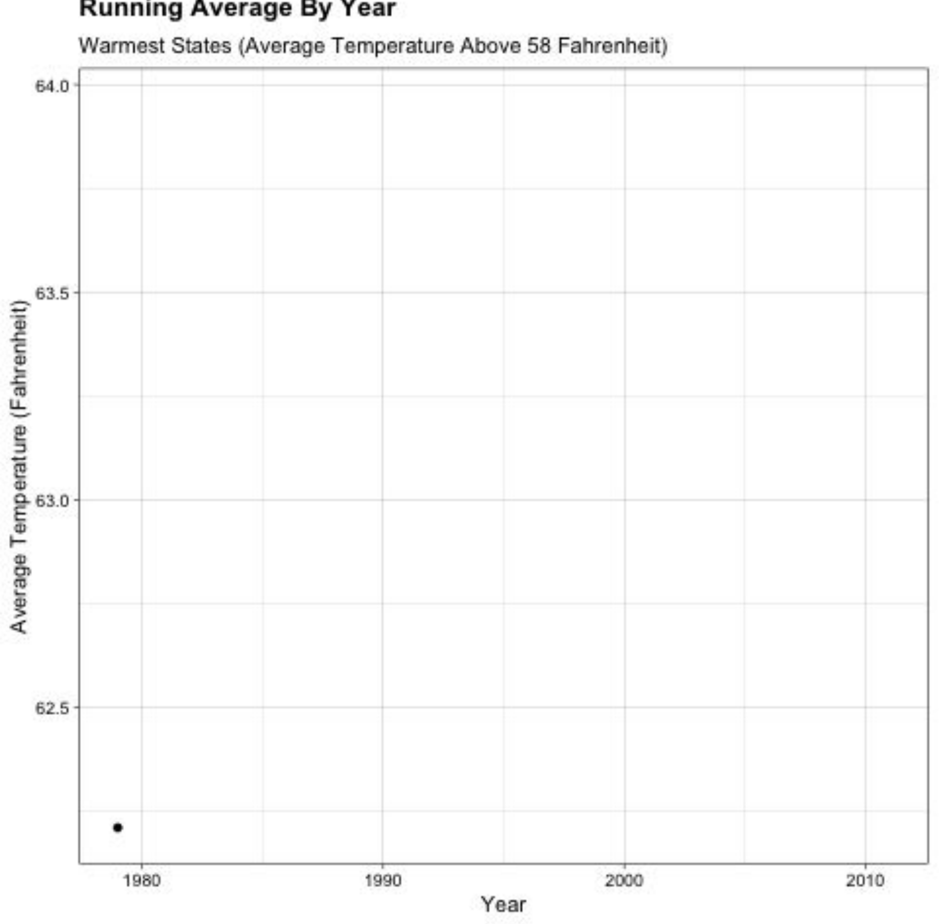
Alabama
Arizona
Arkansas
California
Florida
Georgia
Louisiana
Mississippi
North Carolina
Oklahoma
South Carolina
Texas

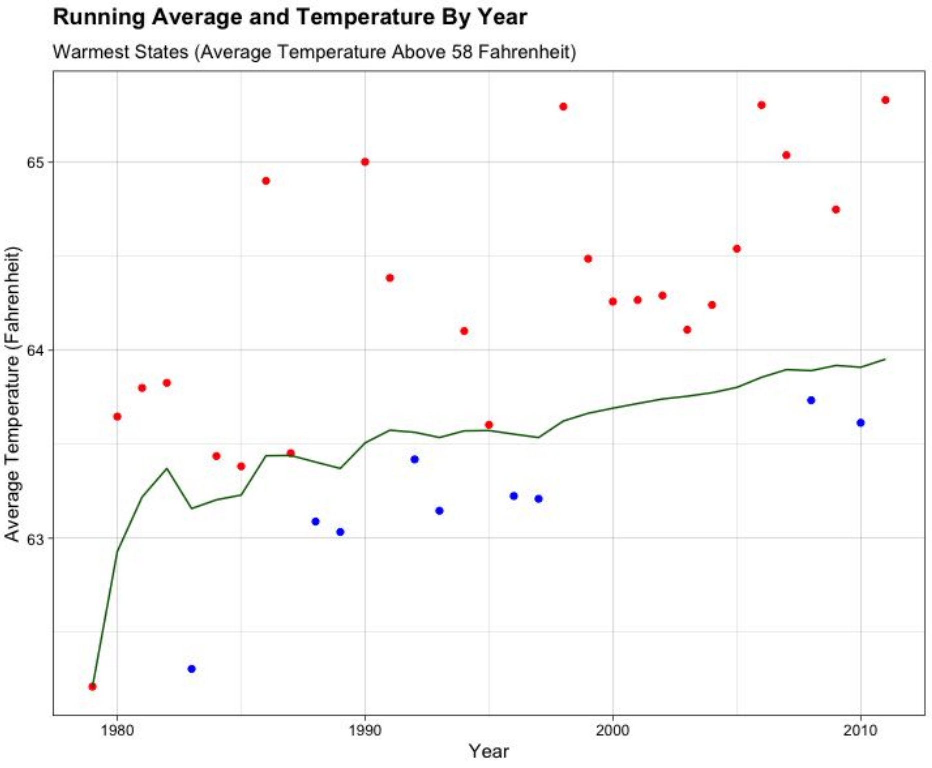




≥ Running Mean

≤ Running Mean





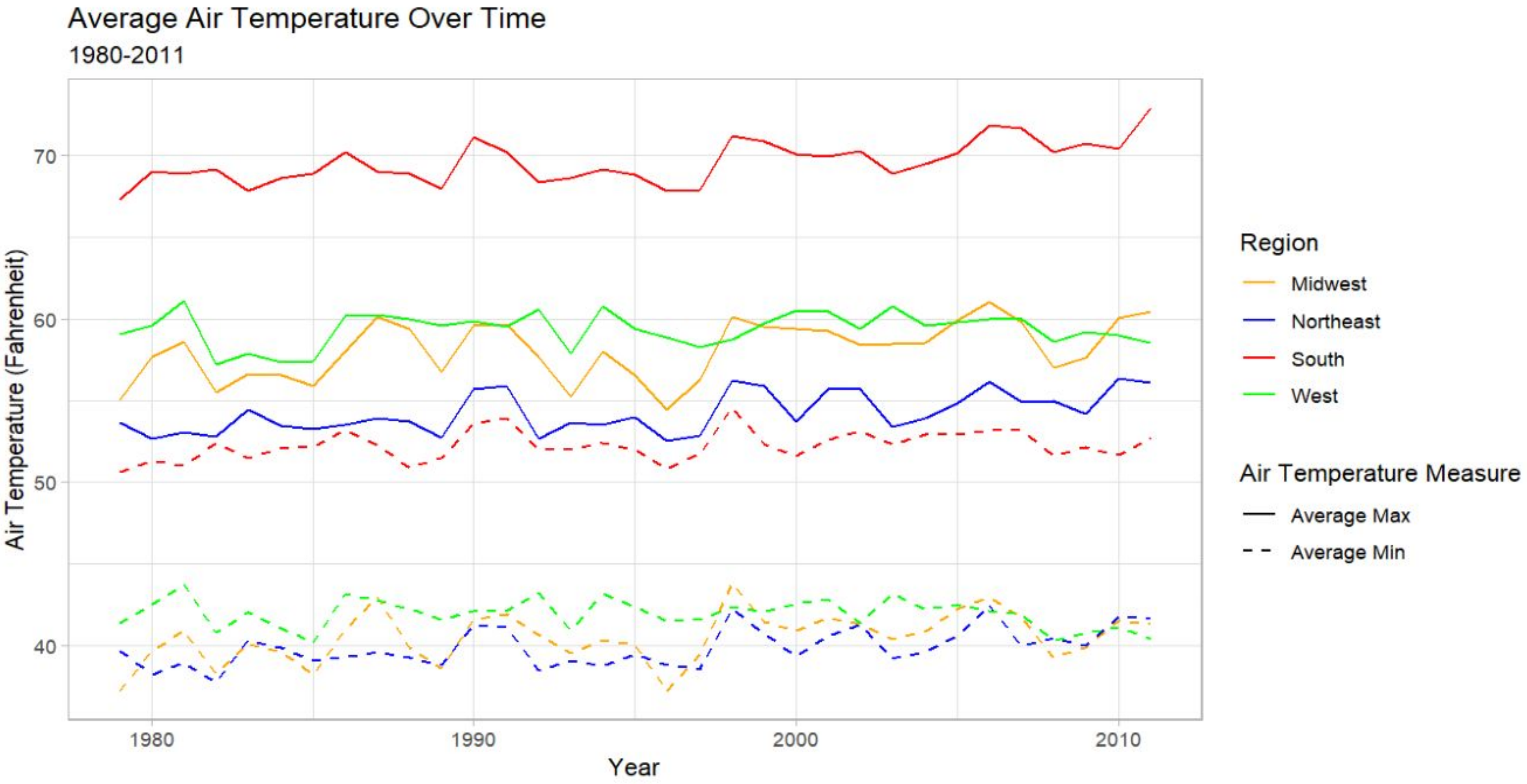
≥ Running Mean
≤ Running Mean

Note: similarity in running avg. plots

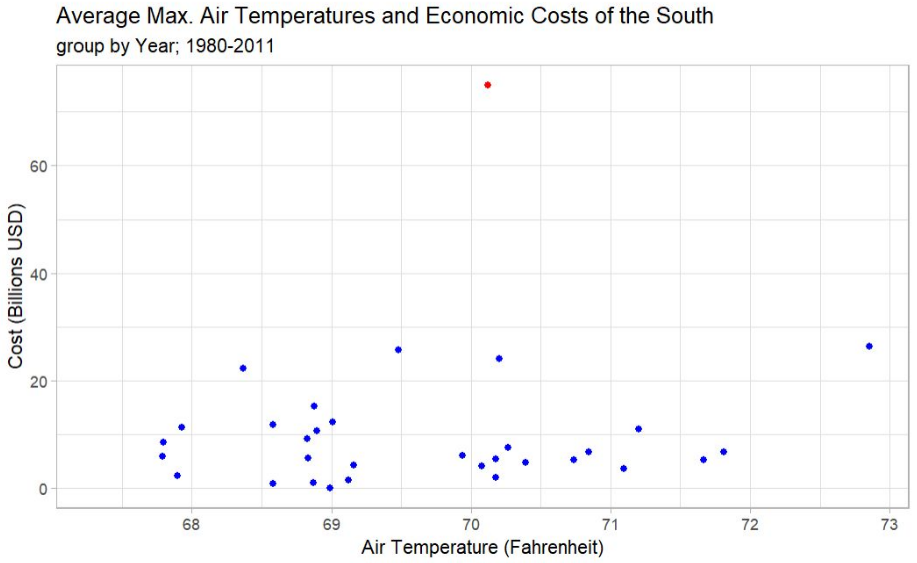
Case Study: The South (Average Temperatures and Economic Costs)

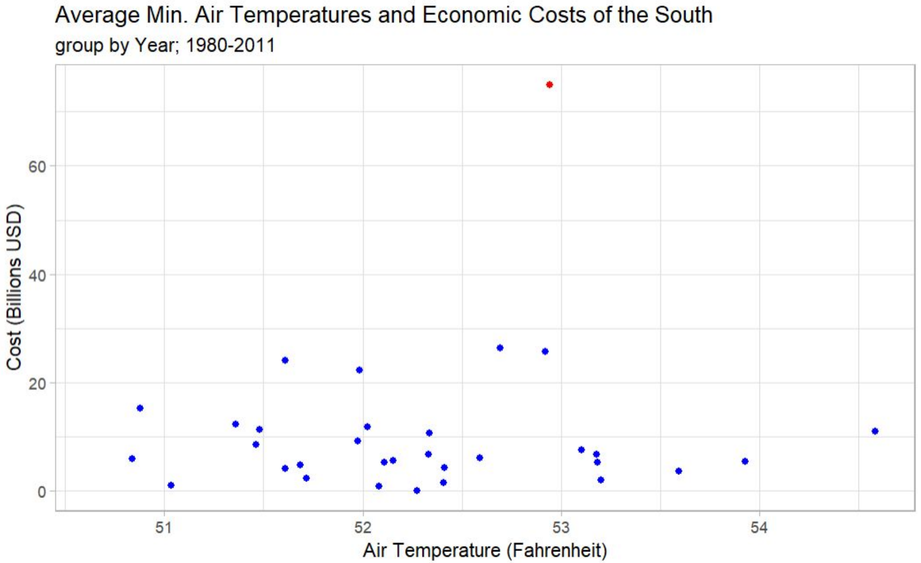
Note: NOAA data begins at 1980, so we ignore the CDC data for 1979

- The South Region contains 16 states (and includes District of Columbia)
- Determine if there's any correlation between air temperature and economic costs from 1980-2011 (32 years)



- Average minimum and maximum air temperatures over time from 1980 to 2011
- Significantly can see the extreme high air temperatures of the south compared to other regions

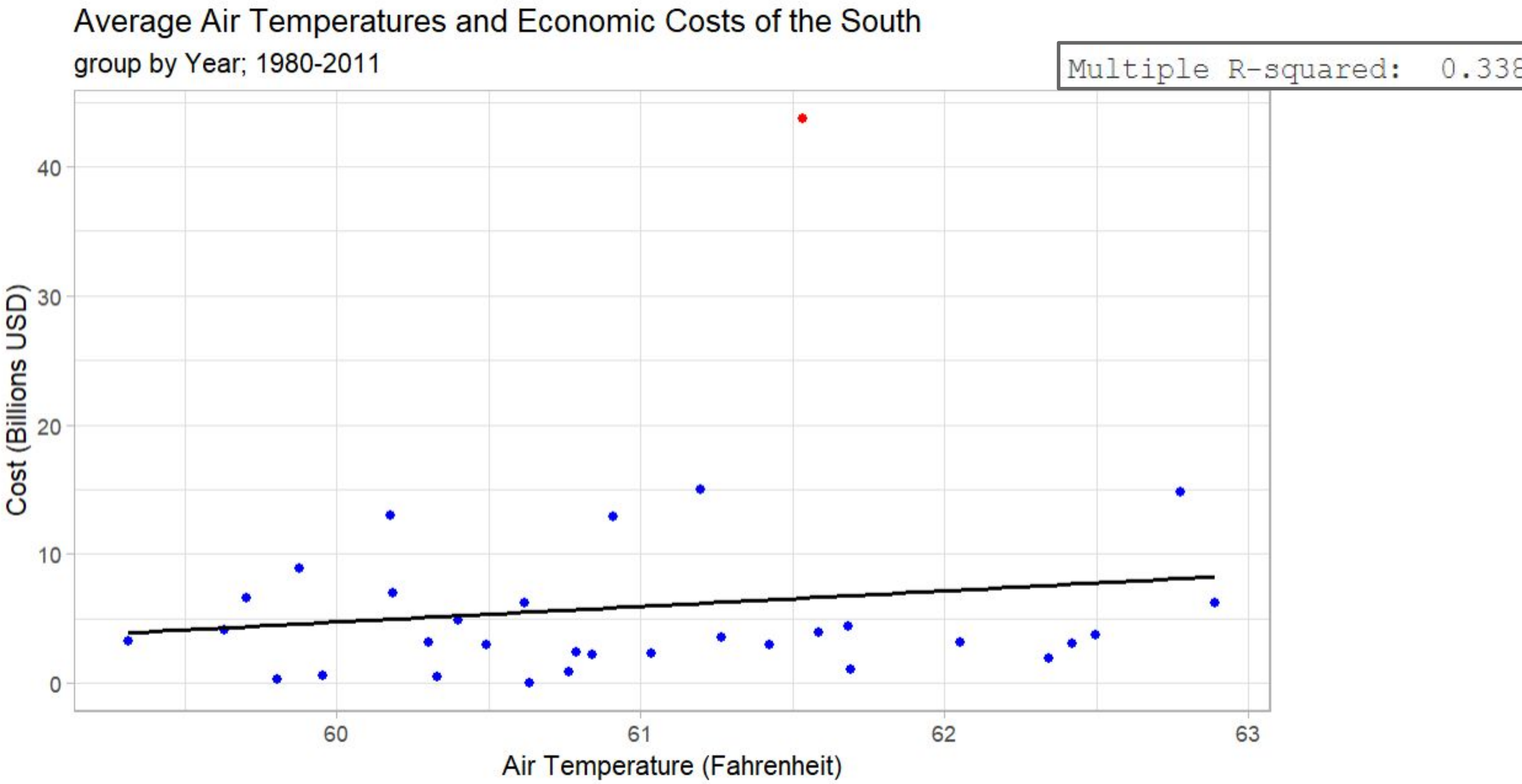




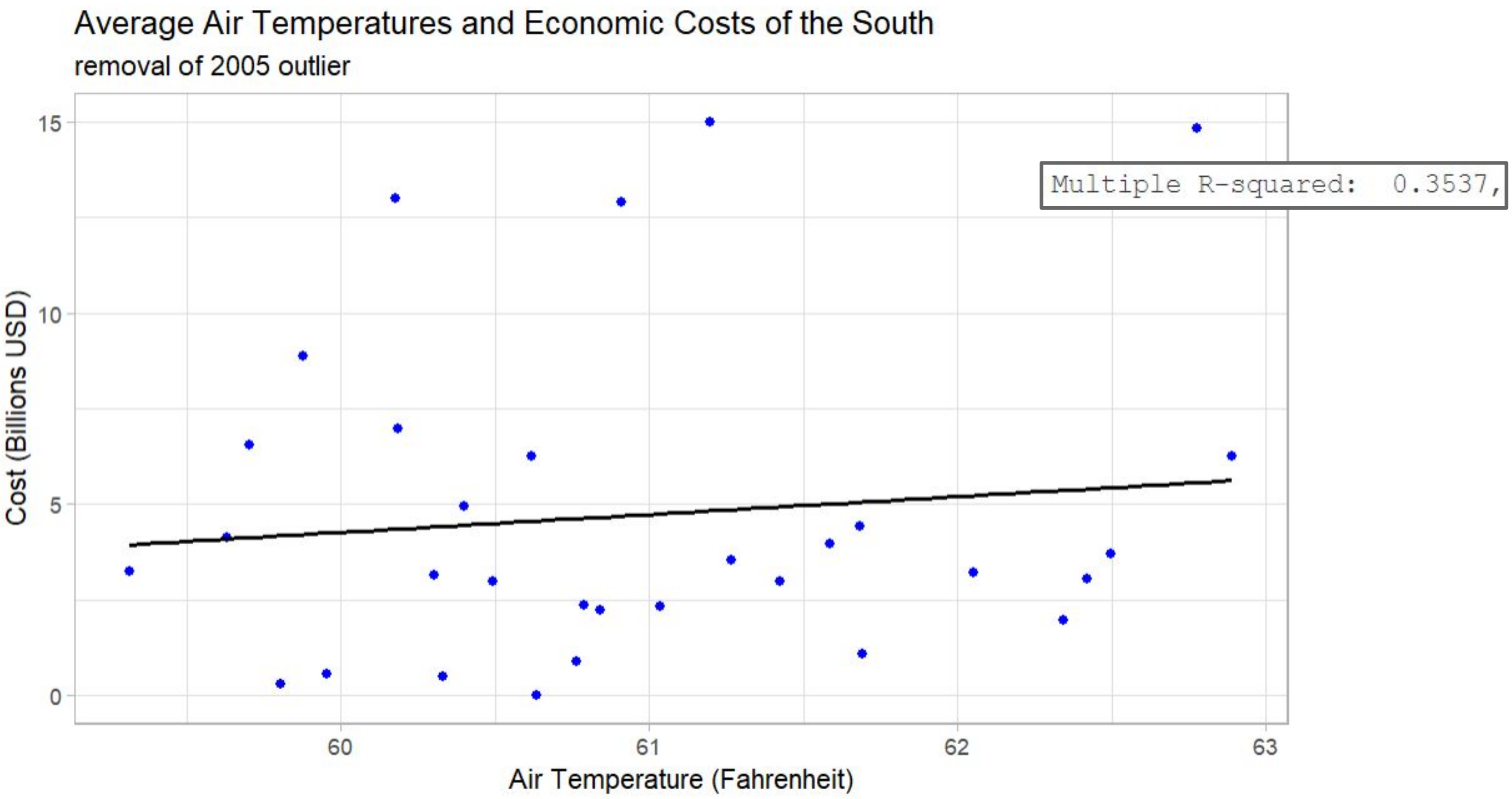
Combine Average Daily Min. and Max. → Average Daily Air Temperature

Average Daily Air Temperature = (Average Daily Min. Air Temperature + Average Daily Max Air Temperature)/ 2

Is there a correlation between Average Daily Air Temperature and Economic Costs of Natural Disasters?



- Based on the statistical summary of the average variables, we output an R-Squared of 0.3381
- The outlier (red point) represents the year 2005. Historically, for natural disasters in 2005 was Hurricane Katrina



Even when removing the significant outlier (2005), the data continues to display a weak correlation (0.3537)

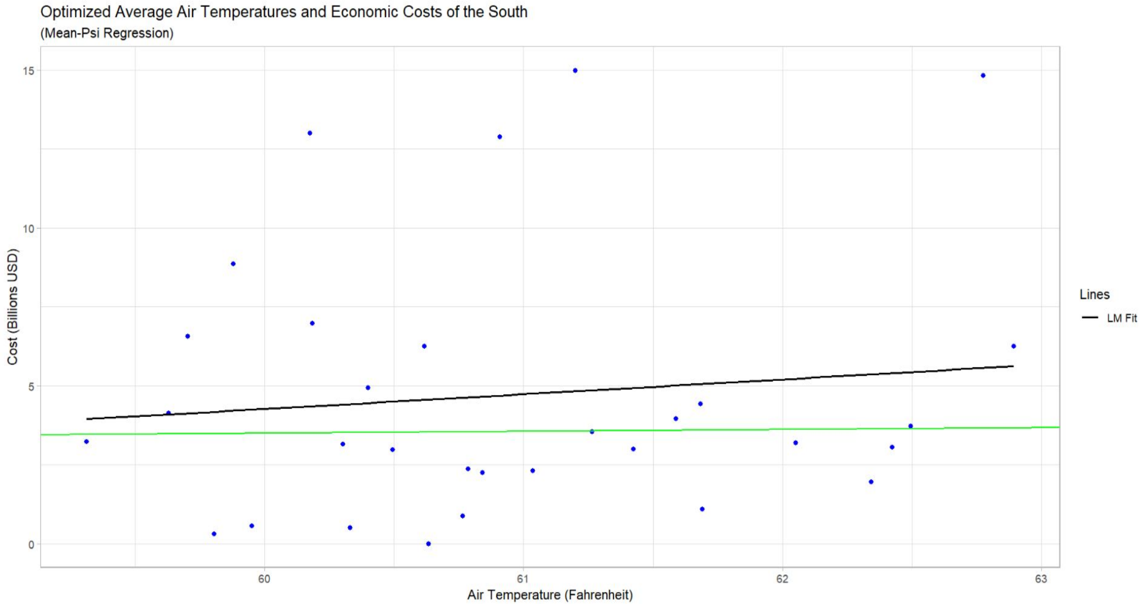
```{r}
psi <- function(x, c = 1) {
  return (
    ifelse(
      abs(x) > c,
        2 * c * abs(x) - c^2,
        x^2
    )
  )
}

mean_psi_regression <- function(b, Y, X, c = 1) {
  return( mean ( psi( Y - ( b[1] + b[2] * X), c ) ) )
}
```

```{r}
initial_params <- c(intercept = 0, slope = 0)
fit3 <- optim(par = initial_params, fn = mean_psi_regression,
              Y = y_data, X = x_data, c = 1, method = "BFGS")

opt_coefs <- fit3$par

cat("Estimated  Intercept:", opt_coefs[1], "\n")
cat("Estimated  Slope:", opt_coefs[2], "\n")
```



- Instead, robust regression based on fitting the psi function and `optim()`
  - `x_data` = average daily air temperature
  - `y_data` = average economic cost (in billions of USD)
  - Converged to estimated intercept of -0.002032337 and estimated slope of 0.05866872
- Conclusion: no true correlation between air temperature and economic costs (Billions of USD)

- Conclusion: no true correlation between air temperature and economic costs
- Possibly due to influence of Congressional Budget Office (CBO) (Investopedia)

Case Study: California Wildfires, Droughts, and Average Temperature

Note: NOAA data begins at 1980, so we ignore the CDC data for 1979

- Due to years of no billion dollar wildfires/droughts, we use a cumulative sum
    - Consistently Increasing ⇒ consistent occurrences
- Compare to cumulative mean of temperature per year (see next slide)

Use of Bootstrapping to Compute Avg. Temperature

- Problem: there are only 12 observations for temperature each year in California (12 months)
- Thus, we cannot be sure of the accuracy of the mean for each year.
- ⇒ Bootstrap each year to estimate the mean temperature per year

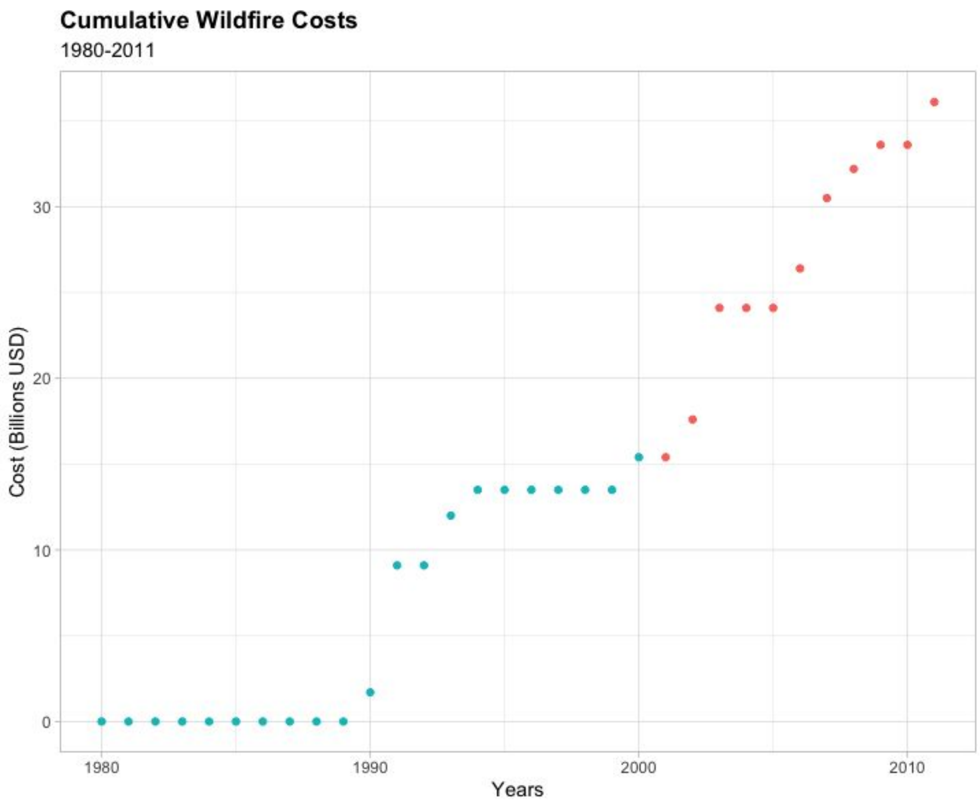
1. Get the data for each year (1980-2011)
    a. 12 observations (January-December)
2. For each year:
    a. Sample the mean of permutations of the dataset (with replacement)

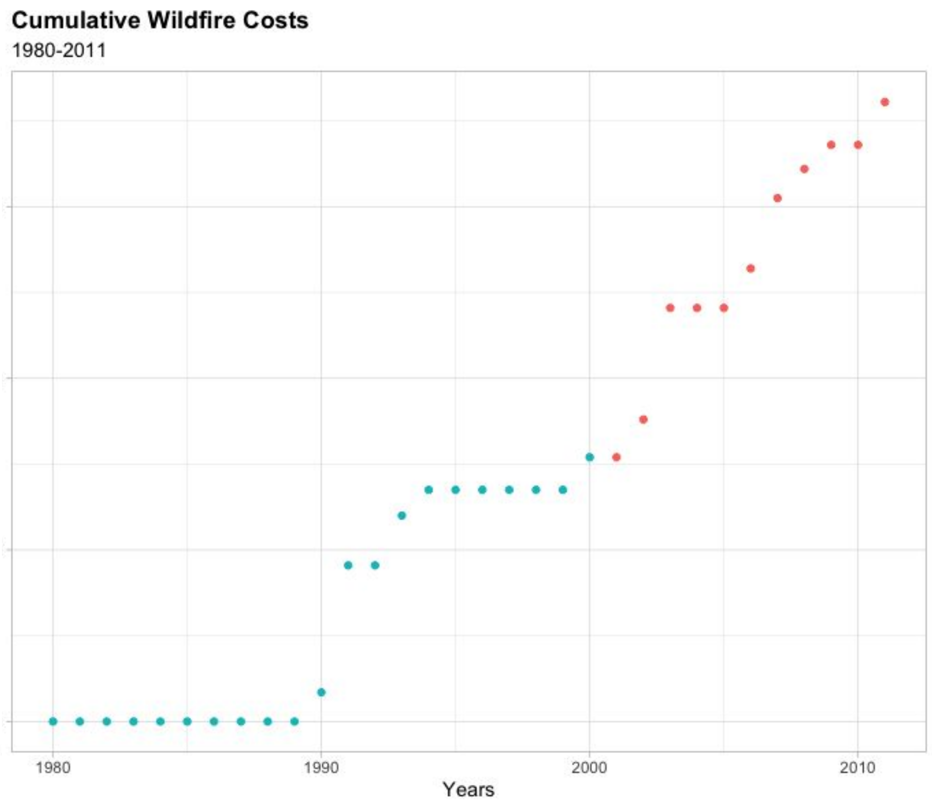
```r
cali_data <- base_data %>%
  filter(State == "California")
df_get <- function(temperature) {
  df_list <- lapply(seq(1980, 2011), function(n) {
    df <- cali_data %>%
      filter(Year == n)
    return(df)
  })
  return(df_list)
}
```

1.

```r
cali_avg_year <- sapply(cali_dfs_avg, function(df){
    S <- 1000
    est_mean <- numeric(S)
    for (n in 1:S) {
      ind <- sample(nrow(df), replace = T)
      list <- df[ind,5]
      est_mean[n] <- mean(list %>% unlist())
      final_mean <- mean(est_mean)
    }
    return(final_mean)
  }
  )
```

2.





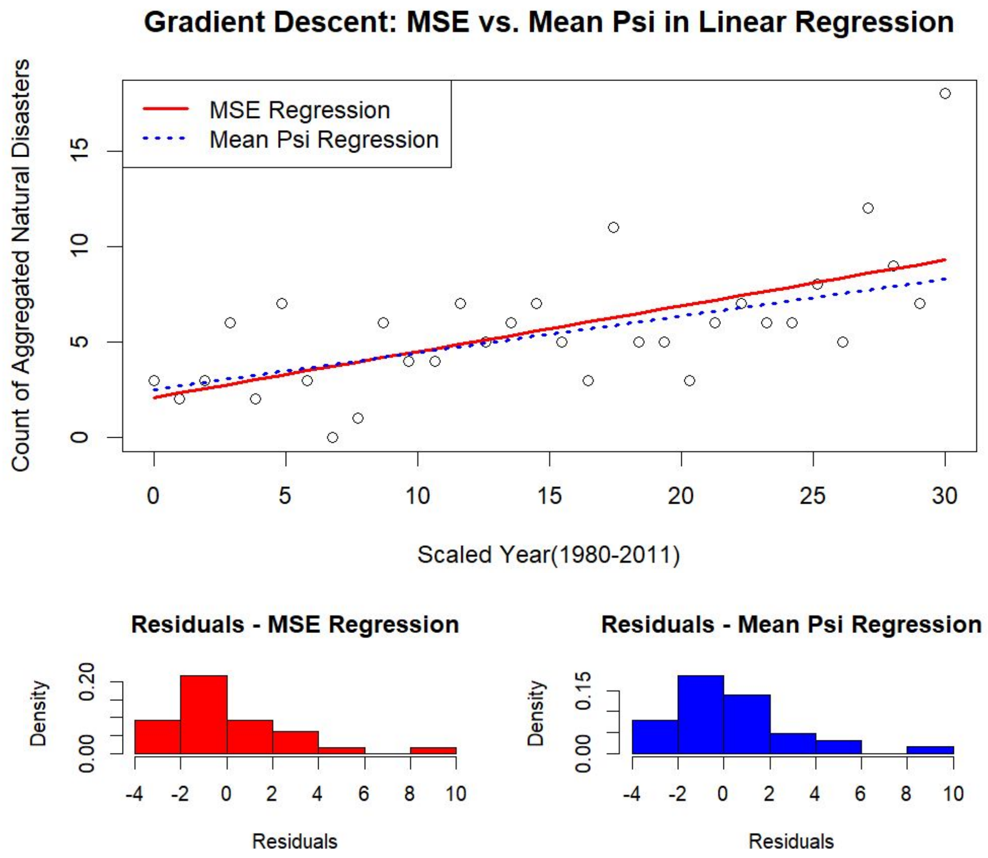
There is a very similar trend between the years 2000 and 2011. Large ‘jumps’ in cumulative sum in 2004, 2007.

Problem:
There is a lot of volatility in the frequency of Natural Disasters. A linear model is likely not the best fit.

$$\mathrm{psi}(x, c) = \begin{cases} x^2 & \text{if } |x| \le c \\ 2c|x| - c^2 & \text{if } |x| > c \end{cases}$$

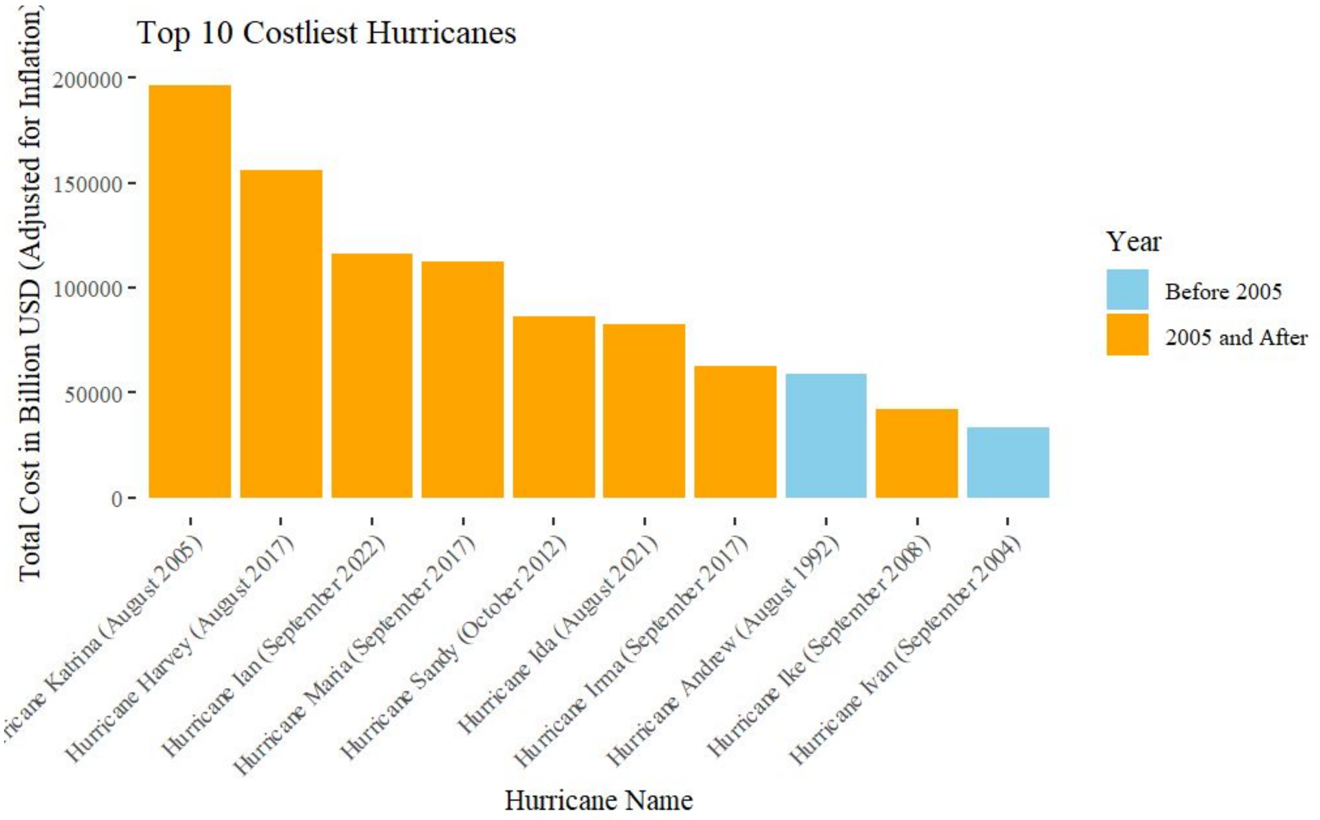
Simple Robust Linear Regression for Frequency of Disasters

- Problem: An outlier in 2011 means that Least Squares Fit is unlikely the best fit for the model
- Robust Linear Regression is more resistant to outliers and therefore can a better model
- ⇒ Run the Gradient Descent Function with a Mean Psi Regression for a better fit



Case Study: Hurricane Katrina

- Note: Average Daily Temperature data ends in 2010 so we ignore the Energy Consumption Data past that year
  -> All Energy Consumption units have been standardized to Billion BTU



Perhaps the most notorious Natural Disaster in modern US history, but could anything have been done to prevent it?

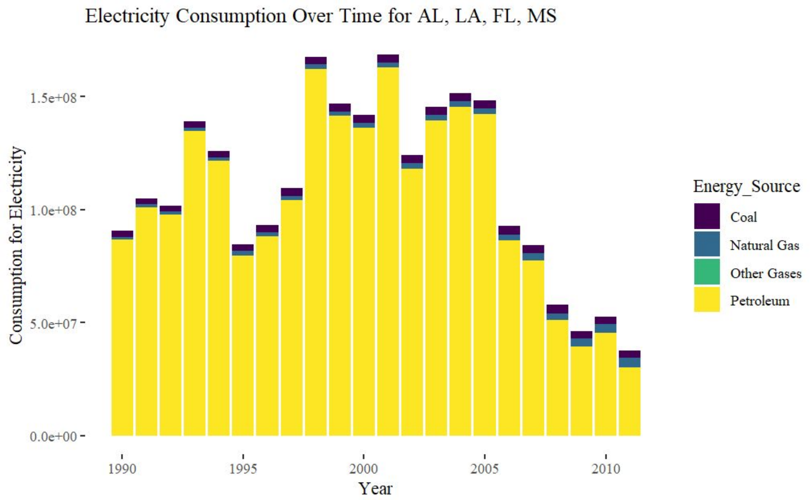
Energy Consumption Data

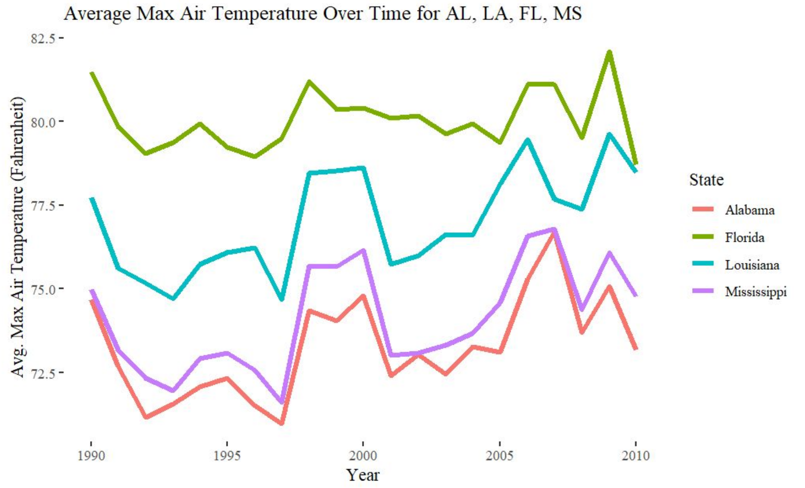
```
'data.frame':	24809 obs. of  11 variables:
 $ YEAR                            : int  1990 1990 1990 1990 1990 1990 1990 1990 1990 1990 ...
 $ STATE                           : chr  "AK" "AK" "AK" "AK" ...
 $ TYPE.OF.PRODUCER                : chr  "Total Electric Power Industry" "Total Electric Power Industry" "Total Electric Power Industry" "Electric Generators, Electric Utilities" ...
 $ ENERGY.SOURCE...............UNITS.: chr  "Coal (Short Tons)" "Petroleum (Barrels)" "Natural Gas (Mcf)" "Coal (Short Tons)" ...
 $ CONSUMPTION.for.ELECTRICITY     : chr  "404,871" "961,837" "42,764,948" "290,182" ...
 $ X                               : logi  NA NA NA NA NA NA ...
 $ X.1                             : logi  NA NA NA NA NA NA ...
 $ X.2                             : logi  NA NA NA NA NA NA ...
 $ X.3                             : logi  NA NA NA NA NA NA ...
 $ X.4                             : logi  NA NA NA NA NA NA ...
 $ X.5                             : logi  NA NA NA NA NA NA ...
```

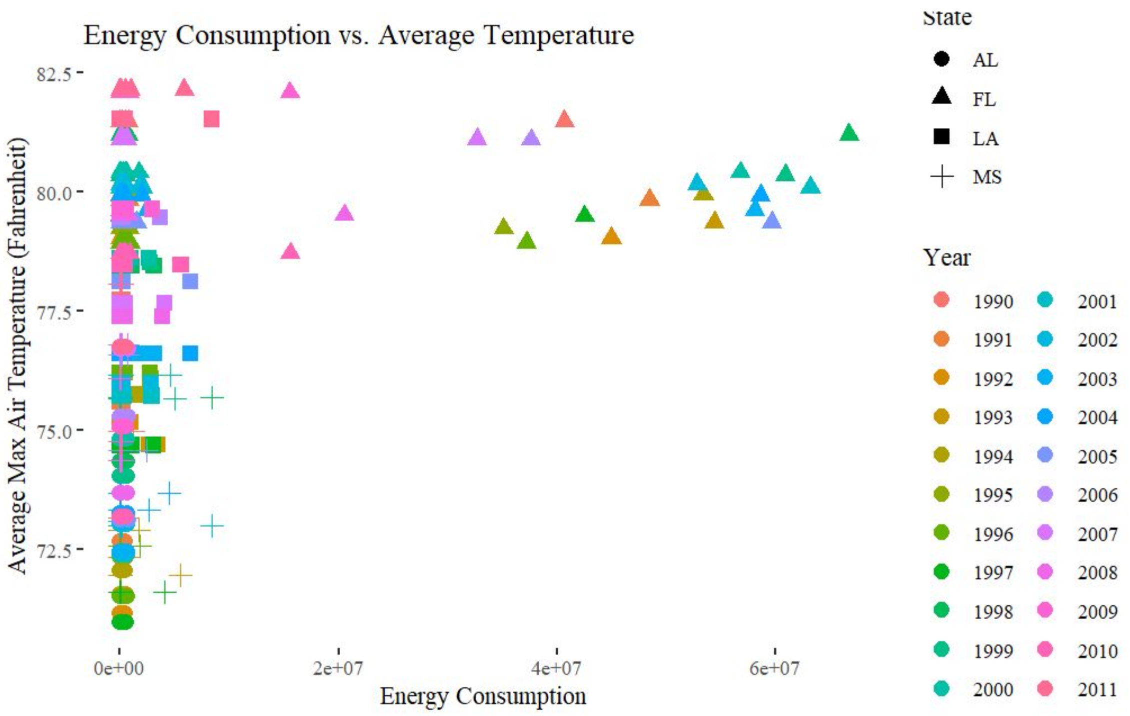
```
'data.frame':	24802 obs. of  5 variables:
 $ YEAR                       : int  1990 1990 1990 1990 1990 1990 1990 1990 1990 1990 ...
 $ STATE                      : chr  "AK" "AK" "AK" "AK" ...
 $ TYPE.OF.PRODUCER           : chr  "Total Electric Power Industry" "Total Electric Power Industry" "Total Electric Power Industry" "Electric Generators, Electric Utilities" ...
 $ Energy_Source              : chr  "Coal" "Petroleum" "Natural Gas" "Coal" ...
 $ CONSUMPTION.for.ELECTRICITY: num  404871 961837 42764948 290182 657706 ...
```

Factors that may have played a part

- Selected 4 states that had the largest impact from Hurricane Katrina
- Peaks and upward trends around 2005 for both Average Max Air Temperature and Electricity Consumption







$$\text{Average Daily Max Temp} = b_0 + b_1 \times \text{Electricity Consumption} + b_2 \times \text{Year} + \epsilon$$

Multiple R-squared: 0.06652,

$$\text{Average Daily Max Temp} = b_0 + b_1 \times \text{Electricity Consumption} + b_2 \times \text{Year} + b_3 \times \text{State} + \epsilon$$

Multiple R-squared: 0.8342,

- Based on the statistical summary of the cumulative variables(Year and Energy Consumption), we output a weak correlation(R-Squared of .06652)
- However, when factoring in state, we can see most of the variation in the data comes when comparing different states

References

Centers for Disease Control and Prevention. (n.d.). *North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011)*. Centers for Disease Control and Prevention. https://wonder.cdc.gov/wonder/help/mcd.html

NCEI.Monitoring.Info@noaa.gov. (n.d.). *Billion-Dollar Weather and Climate Disasters | National Centers for Environmental Information (NCEI)*. Www.ncei.noaa.gov. https://www.ncei.noaa.gov/access/billions/time-series/US

The Financial Effects of a Natural Disaster. (n.d.). Investopedia. https://www.investopedia.com/financial-edge/0311/the-financial-effects-of-a-natural-disaster.aspx#:~:text=According%20to%20a%202019%20Congressional%20Budget%20Office%20%28CBO%29