DSI_Final_Project

Emma Eichelman, Rasmus Jonsson, Colleen Mattingly

1/15/2021

Introduction

Climate change has become one of the most pressing issues in recent years, impacting coastal cities and minority communities at a faster rate than other groups. The climate is changing, forced out of the range of the past million years by levels of carbon dioxide and other greenhouse gases not seen in the Earth's atmosphere for a very, very long time. Lacking action by the world's nations, it is clear that the planet will be warmer, the sea level will rise, and patterns of rainfall will change. But the future is also partly uncertain, there is considerable uncertainty about how we will arrive at that different climate. The public opinion regarding climate change is very divided and differs depending on factors such as age, location, religious beliefs, etc. Therefore, the main purpose of this student analysis is to determine that if you live in a state that has encountered a natural emergency, then you would be more likely to believe in global warming. In addition to geographic location of survey respondents, we looked at other factors that could contribute to the belief of global warming. Studies show that climate change has a disproportionate impact on minorities. The analysis/report also suggests that minorities are more heavily impacted by climate change and natural disasters. After carefully considering different options, the student group decided to base the analysis of the following research question: "does a person's location (based on state within the United States) influence their feelings towards climate change?". In addition to the research question developed, the null hypothesis established by the student group was "the presence of a state emergency will increase respondent's belief in climate change".

Methodology

Raw Data & Data Cleaning

Both public opinion survey data and data around natural disasters in the United States were used for the purpose of this analysis. The public opinion survey data was obtained from openICPSR, an open source repository for social, behavioral, and health sciences research data. The National Surveys on Energy and Environment [United States] for years 2010, 2014, and 2018 were used for this analysis. The National Survey on Energy and Environment (NSEE) is a biannual opinion survey on energy and climate policy that is conducted through randomized phone calls of United States (U.S.) residents over the age of 18 years. A series of demographic questions were asked regarding the participants race, age, location, political affiliation, etc. along with questions regarding their agreement with various environmental and energy policies and general sentiment towards climate change. Respondents are not required to answer so for the purpose of this analysis only certain questions and their responses were included. Open FEMA data was used to identify natural disasters by state, year over year. Natural disasters are categorized as major adverse events resulting from the natural processes of the Earth. FEMA identifies natural disasters as fires, severe storms, earthquakes, floods, snow, tornadoes, coastal storms, landslide, ice storms, tsunamis, etc. NSEE data: Prior to formally testing our hypothesis, the datasets for the 2010, 2014, and 2018 NSEE surveys needed to be cleaned so they could be compiled into one final dataset for analysis. Any responses containing a 98 ("Not Sure") were removed and responses with code 99 ("No Response") were recorded to NA for clarity. For the purpose of this analysis, only columns containing information for race, age, education, income, state location of the respondent, and response to the question "Is there solid evidence that the average temperature on earth has been getting

warmer over the past four decades? (Yes/No)", were included in the three final datasets. This question is intended to gauge a respondents belief in climate change, which was why we chose these responses to analyze. FEMA data: For our analysis we filtered out states that have experienced a fire, severe storm, or hurricane in the year 2009, 2013, and 2017 and recoded them as factor "1". The reason the years previous to the survey data were used was to make sure there was not a misalignment with timing of potential disasters and the collection of the survey responses. We chose to narrow the results by severe storm, hurricanes, and fire as those are three natural disasters. Any other state that did not experience any of those three natural disasters were given a "0". We created new variables for each year that covered the states that were combined into the final dataframe. Our state emergency variable sample included the states that experienced an emergency in the prior year to the survey to account for timing.

Analysis

Our analysis began with a proportion test between our 2010 and 2018 NSEE data to observe the proportional differences. In order to determine if the location of a respondent, and therefore experience with natural disasters, impacts their belief in climate change, the general logistic model (GLM) was used.

Preliminary Proportion Tests

Proportion of respondents living in a state that has experienced a natural disaster to their response on belief in climate change

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  table(df.compfinal$gw_belief, df.compfinal$state_emergency)
## X-squared = 0.22407, df = 1, p-value = 0.636
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.03952325  0.06460203
## sample estimates:
##    prop 1    prop 2
## 0.6237705 0.6112311
```

Proportion Test Summary

We ran a two-sample proportion test using our complete dataframe covering the years from 2010-2018 to show the proportion of people living in states that encountered a natural disaster and their response to the question on whether they believe in climate change. Our results from the two-sample proportion test did not provide a statistically significant outcome because the p value is larger than our alpha. Using this prop test as a a basis, we dug deeper into the data to run logistical models year by year.

Binary Logistic Regressions

Logistic Regression of Belief in Climate Change in 2010

```
logit_b <- glm(gw_belief ~ demog_race + demog_age_list + demog_edu + demog_income + state_emergency, fa
summary(logit_b)
```

```
##
## Call:
## glm(formula = gw_belief ~ demog_race + demog_age_list + demog_edu +
##     demog_income + state_emergency, family = binomial, data = df.beg_final)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1791  -0.9841  -0.8897   1.3485   2.2751
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.170036   0.496691  -2.356  0.01849 *
## demog_raceAfrican American -1.744205   0.616917  -2.827  0.00469 **
## demog_raceAsian             -0.828541   1.132565  -0.732  0.46444
## demog_raceLatino            -0.543645   0.533251  -1.019  0.30797
## demog_raceMixed race        -1.374022   0.769761  -1.785  0.07426 .
## demog_age_list               0.139153   0.114173   1.219  0.22292
## demog_edu                    0.009481   0.094653   0.100  0.92021
## demog_income                 0.053514   0.068349   0.783  0.43366
## state_emergency              0.119343   0.190382   0.627  0.53075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 663.97  on 511  degrees of freedom
## Residual deviance: 641.97  on 503  degrees of freedom
## AIC: 659.97
##
## Number of Fisher Scoring iterations: 4
```

2010 Summary

The 2010 dataframe started with 751 observations with 51 variables, however in the cleaning process, it was reduced to 557 observations with 7 variables. The data cleaning removed the "not sure" coded as a 98 or NA codes. We then ran a glm to examine the likelihood of belief in global warming based on the factor of race, age, highest education level obtained, income, and whether their state had experienced a natural emergency during 2009. For 2010, we can interpret these results showing the estimated coefficients on the likelihood of belief in climate change, a binary and categorical response. In comparing these results, we cannot draw official conclusions because we did not find a statistically significant result. Based on these initial results, we cannot reject our null hypothesis.

Logistic Regression of Belief in Climate Change in 2014

```
logit_d <- glm(gw_belief ~ demog_race + demog_age_list + demog_edu + demog_income + state_emergency, fa
summary(logit_d)
```

```
##
## Call:
## glm(formula = gw_belief ~ demog_race + demog_age_list + demog_edu +
##     demog_income + state_emergency, family = binomial, data = df.mid_final)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3212  -0.9101  -0.7825   1.3103   1.8798
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                -0.27612    0.45506  -0.607  0.54399
## demog_raceAfrican American -0.29957    0.33847  -0.885  0.37612
## demog_raceAsian            -0.22635    0.68730  -0.329  0.74190
## demog_raceLatino            0.43877    0.38545   1.138  0.25499
## demog_raceMixed race       -0.63366    0.52592  -1.205  0.22826
## demog_age_list              0.21002    0.09854   2.131  0.03306 *
## demog_edu                  -0.27853    0.10385  -2.682  0.00732 **
## demog_income                0.02323    0.06772   0.343  0.73154
## state_emergency            -0.33723    0.19441  -1.735  0.08281 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 700.81  on 547  degrees of freedom
## Residual deviance: 679.48  on 539  degrees of freedom
## AIC: 697.48
##
## Number of Fisher Scoring iterations: 4
```

2014 Summary

In 2014, the dataset started with 798 observations with 70 separate variables which later were reduced to 494 observations with 7 variables. As always when a big segment of a respondent pool is removed, the answers might be more or less skewed. For 2014, we interpreted that while the overall age increased, the belief in climate change (gw_belief) decreased. In addition, the state emergency (state_emergency) turned out to be somewhat statistically significant even though the related data was relatively slim which can make it difficult to make any major conclusions.

Logistic Regression of Belief in Climate Change in 2018

```
logit_c <- glm(gw_belief ~ demog_race + demog_age_list + demog_edu + demog_income + state_emergency, fa
summary(logit_c)
```

```
##
## Call:
## glm(formula = gw_belief ~ demog_race + demog_age_list + demog_edu +
##     demog_income + state_emergency, family = binomial, data = df_final)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.91338  -0.67292  -0.51511  -0.00017   2.42506
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -2.77857    0.51166  -5.431 5.62e-08 ***
## demog_raceAfrican American -17.01964  897.21694  -0.019   0.9849
## demog_raceAsian            -16.92468 1713.28458  -0.010   0.9921
## demog_raceLatino            -0.69427    0.49800  -1.394   0.1633
## demog_raceMixed race        -0.01486    0.37439  -0.040   0.9683
## demog_age_list               0.41577    0.11500   3.615   0.0003 ***
## demog_edu                   -0.06242    0.10683  -0.584   0.5590
## demog_income                 0.07151    0.07141   1.001   0.3166
## state_emergency              0.21514    0.23711   0.907   0.3642
```

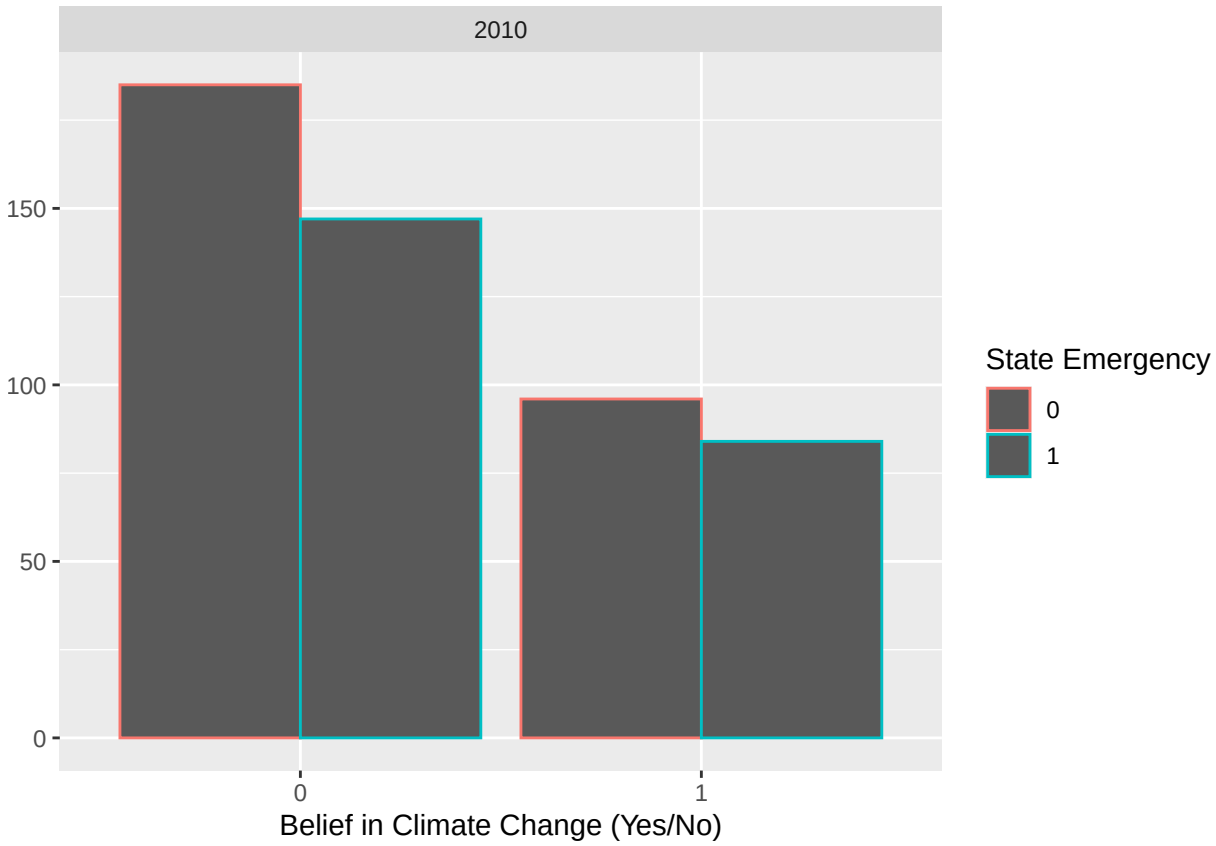
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 542.22  on 622  degrees of freedom
## Residual deviance: 497.65  on 614  degrees of freedom
## AIC: 515.65
##
## Number of Fisher Scoring iterations: 17
```

2018 Summary

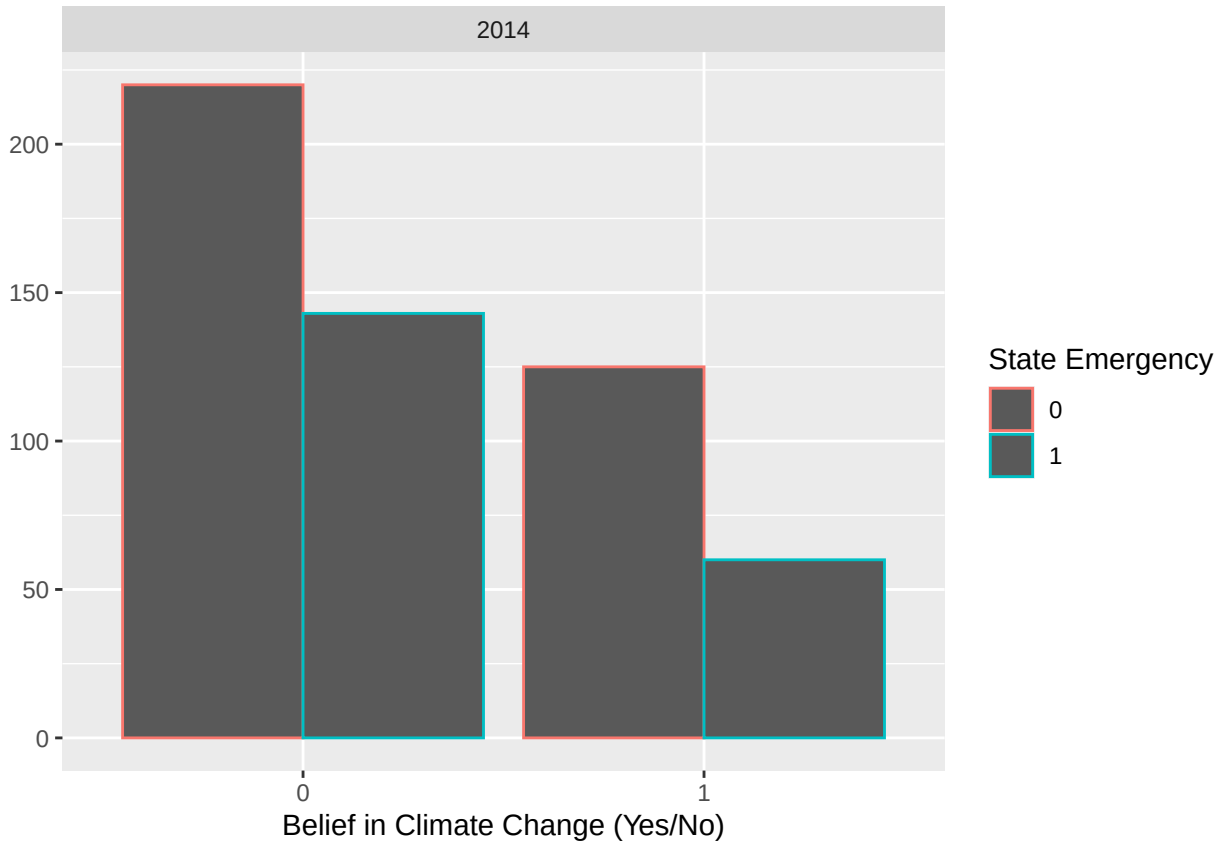
In 2018, the dataset started with 751 observations and 51 variables. After removal of variables not within the analysis and any non-responses or responses of "not sure," the total observations were reduced to 557 with 7 variables. This significant loss of observations could have had a negative impact on the final analysis of the data for this year. All respondents that identified as African-American and Asian responded "no" to the question on belief in climate change. This gave a mean of 0 for both groups and greatly skewed the standard error as seen above. In 2018, age continued to be the most significant factor with belief in climate change increasing as age increases. State emergency had a positive correlation in 2018, but was not statistically significant so no conclusions could be drawn on how an increase in natural disasters impacts belief in climate change.

Graphs

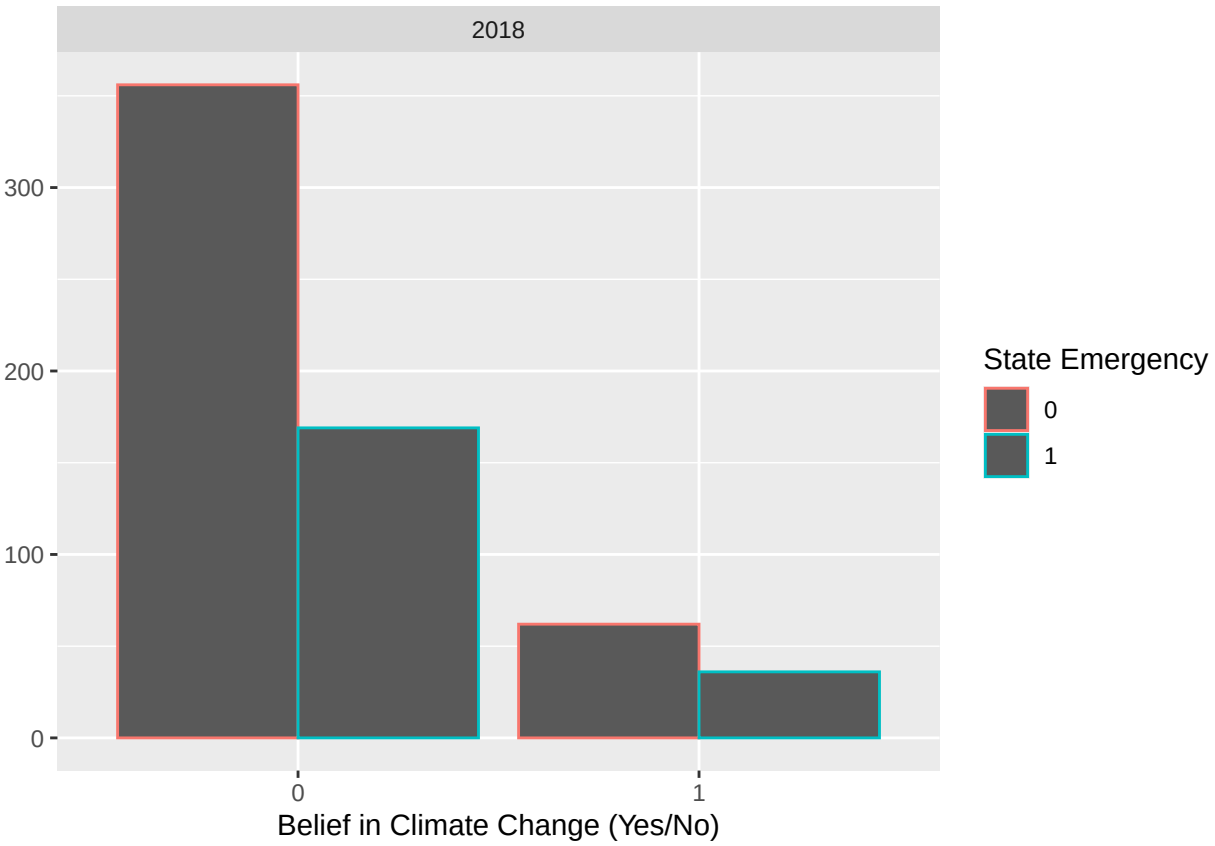
```
ggplot(df.beg_final, aes(x = gw_belief %>% as.factor)) +
  geom_bar(aes(color = state_emergency %>% as.factor), position='dodge') +
  labs(x = "Belief in Climate Change (Yes/No)", y = NULL, color = "State Emergency") +
  facet_wrap(~Year)
```




```
ggplot(df.mid_final, aes(x = gw_belief %>% as.factor)) +
  geom_bar(aes(color = state_emergency %>% as.factor), position='dodge') +
  labs(x = "Belief in Climate Change (Yes/No)", y = NULL, color = "State Emergency") +
  facet_wrap(~Year)
```




```
ggplot(df_final, aes(x = gw_belief %>% as.factor)) +
  geom_bar(aes(color = state_emergency %>% as.factor), position='dodge') +
  labs(x = "Belief in Climate Change (Yes/No)", y = NULL, color = "State Emergency") +
  facet_wrap(~Year)
```



Graph Discussions

Creating graphs when the variables were primarily categorical and binary was challenging. The above graphs show how responses to gw_belief have changed year over year based on state emergency. These graphs support the results of the statistical analysis, discussed below, and accept the null hypothesis.

Discussion

In the end, we failed to reject the null hypothesis in years 2010, 2014, and 2018. This failure to reject the null hypothesis could be in part due to the failure of the analysis to differentiate one disaster per year versus multiple disasters per year. The significance of state emergency may have increased because there were more total disasters, which impacted more people's opinions, but this cannot be determined. There are many other variables that can influence individuals' attitudes towards climate change that were not included in this analysis. There are two major limitations in this analysis that could be addressed in future research. The first limitation of this study was time constraints. More variables could have been added into the final analysis or additional tests performed given more time. The second limitation worth mentioning is unanswered questions within the data pulled for the analysis. The large number of NAs meant that a lot of respondents had to be removed from the original data set and this could potentially really skew the results. Finally, people's beliefs on climate change can be influenced by many external factors not included in this analysis, including political beliefs, opinions of family members, etc.

Bibliography Christopher Borick, M. C. I. of P. O., Sarah Mills, U. of M., & Barry Rabe, U. of M. (2019). National Surveys on Energy and Environment [United States] [Data set]. Inter-university Consortium for Political and Social Research (ICPSR). https://doi.org/10.3886/E100167V19 Milfont, T. L., Evans, L., Sibley, C. G., Ries, J., & Cunningham, A. (2014). Proximity to Coast Is Linked to Climate Change Belief. PLoS ONE, 9(7). https://doi.org/10.1371/journal.pone.0103180 OpenFEMA Dataset: FEMA Web Disaster

Declarations—V1 | FEMA.gov. (n.d.). Retrieved January 15, 2021, from https://www.fema.gov/openfema-data-page/fema-web-disaster-declarations-v1 Race, Ethnicity and Public Responses to Climate Change. (n.d.). Yale Program on Climate Change Communication. Retrieved January 15, 2021, from https://climatecommunication.yale.edu/publications/race-ethnicity-and-public-responses-to-climate-change/