Data-driven Research

Are the Predictors of Climate Unawareness and Climate Denial Different?

Minerva University

NS125: Research Methods

Prof. P. Zoogman

March 6, 2023

Are the Predictors of Climate Unawareness and Climate Denial Different?

Introduction

Research (Kahan et al., 2011) indicates that deeply rooted climate change denial is resistant to change by factual information, suggesting that efforts might be better focused on the uninformed or ambivalent population. Research on predictors fo climate unawareness can be valuable for effective policy decesions.

Data Prep

To analyze the Howe et al. (2015) dataset I started with subsetting the data into the state level and county. Then I split the GeoName column into "State Name" and "County Name" columns in the Sateslevel subset[1] of the original data in order to combine it with the Geographic map data of the "maps" library. These steps also ensured that my data was tidy. I also created a new column called DNhappening to record the estimates of the percentage of people who do not know global warming is happening[2] in both the subsets of data. I also created columns to store the relative differences from the national average for climate-aware, unaware, and denier estimates[3]. Moreover, while presenting data I changed the column names so they are easily understandable by the audience.

Data Analysis

Research shows there is variation in climate awareness at the State and County level but I wanted to see if that is true for climate unawareness and climate denial as well. I created maps of the contiguous United States for climate aware, unaware, and denial estimates to visually inspect the differences between variations. From Fig. 1 we can see the variation in the percentage of climate unawareness and climate denier population is almost opposite of each other, while this relationship is not that clear for a percentage of climate unaware population.

[1] #dataprep: I have explained my process and reasoning for the data preparation and how that helped me with my data analysis. For ex, by subsetting the data at state level I was able to create the US maps and make sure the data is tidy and by creating the new column for climate unaware population I was able to analyze their variation at the state and county level.

[2] This estimate was calculated by subtracting the sum of the percentage of the climate aware and climate deniers from 100.

[3] This was calculated by subtracting the original values of the estimates from the national average values

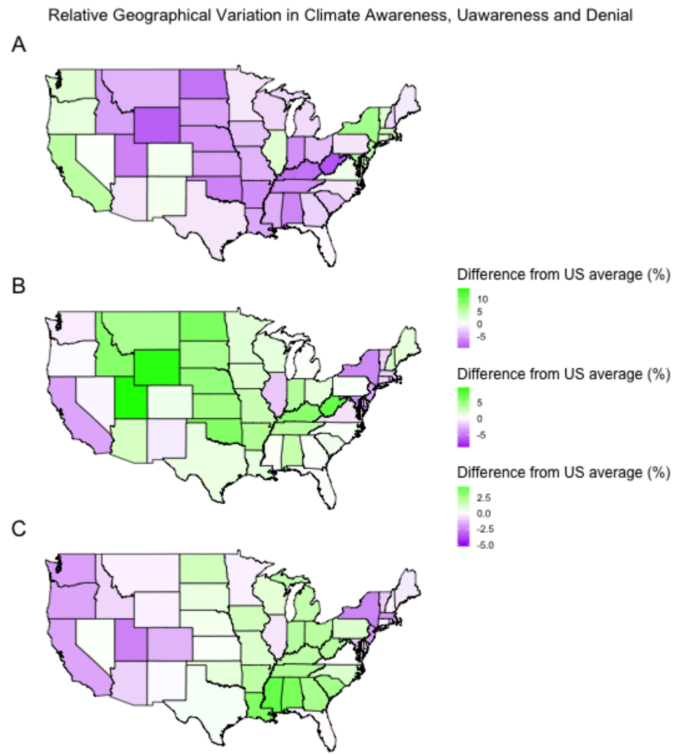

Figure 1: *Geographic variation in the estimated percentage of American adults who a) think that global warming is happening; b) think that global warming is not happening; c) do not know if global warming is happening. Note that these are relative differences from the national average to help comparisons between states[4].*

I then created a box plot to visually inspect the variation in the percentage of the climate-unaware population at the State and County level[5]. From Fig. 2 it is clear that the spread of data is greater at the County level as compared to the State level.

[4] Hawaii and Alaska are not included in the map due to technical reasons

[5] Initially I wanted to create the US map at the county level to inspect this variation but was not able to do so because of high loading time R.

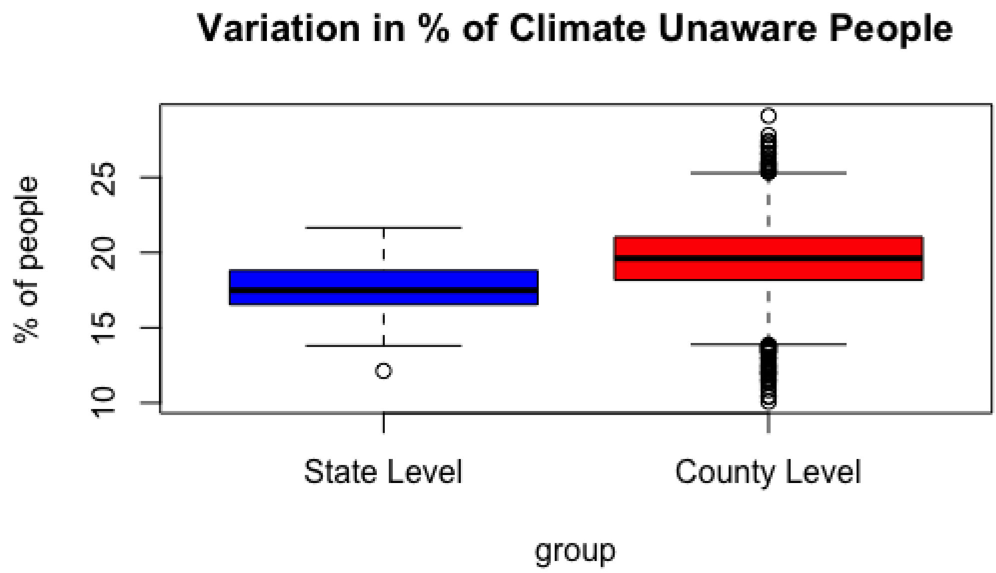

Figure 2: *Difference in variation between the percentage of climate unaware population at the State and County level.*

Looking at that variation at the county level we can also see that there are a number of outliers indicating that there are counties with significantly different high and low percentages of climate-unaware populations. The difference between outliers is as high as ~20 percentage points which is a big difference given this gives the potential percentage of population that can be persuaded to become climate aware.

The observation of outliers is interesting because research on these outliers can help us understand what leads to high and low percentages of climate unawareness at the county level. Before thinking about that, I wanted to analyze the relationship between percentage of climate unaware and climate denier population at the county level to see if there is a correlation. I created a scatterplot and calculated the R^2 value using the pearson's "r" to analyse the relationship. From the Fig. 3 it is clear that the relationship between the two variables is positive but very weak. The R^2 of ~ 0.13 implies that only 13% of the variance in the percentage of climate unaware population can be explained from the percentage of climate denier population.[6]

[6] #dataanalysis: I used the best practices for EDA, asking myself question about the data like "how does climate unaware varies from the national average?", "Is the variation different at the county level?" and "does climate unawareness percentage vary with climate denial percentage at the county"? I went have also explained how I moved from one step to another iteratively.

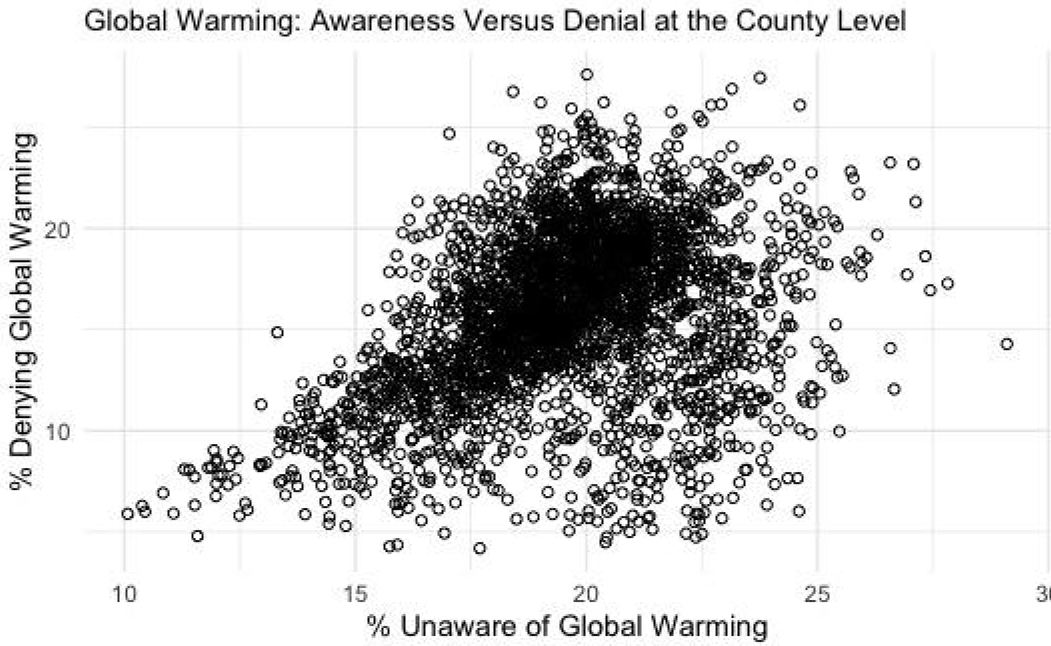

Figure 3: *Scatter plot showcasing the correlation between the percentage of climate-unaware population and the percentage of climate deniers at the county level. From the plot we can see a very weak positive relationship between the two variables. The R^2 value of the correlation between the two variables is ~ 0.13.*[7]

Proposed Methodology

The analysis stemming from Figure 3 reveals a lack of strong correlation between the levels of climate unawareness and climate denial across counties. This observation leads to the hypothesis that the top predictor for climate denial and climate unawareness at the county level might differ. Prior studies have pointed to education as a key determinant in climate awareness (Lee et al, 2015), suggesting that individuals with higher education levels tend to acknowledge climate change due to their exposure to relevant topics, scientific principles, and critical thinking skills developed through higher education. Conversely, political ideology, particularly the proportion of individuals voting for the Republican party, is believed to significantly influence climate denial (McCright et al., 2016). This is attributed to partisan biases, including consumption of conservative media linked with climate skepticism(Ash et al., 2023)[8], and communal cultural pressures.

This research aims to verify these assumptions by examining county-level socioeconomic data from 2013, focusing on variables such as education levels (specifically the percentage of individuals holding at least a bachelor's degree), political affiliations (percentage of republican

[7] #dataviz: I created clear diagrams with appropriate titles, axes, legends and captions that best reflect the message of each visualization.
[8] #plausibilty: I have mentioned multiple plausible assumptions about the basis of my hypothesis and also provided evidence for the assumptions wherever necessery or available.

voters), geographic location (urban vs. rural), and internet access. An ensemble random forest model, similar to that used in Lee et al. (2015), will be employed to identify the leading predictors of climate change unawareness and denial for each county.[9]

Several outcomes are possible from this study. Should the primary predictors align with the hypothesis, it would bolster the argument that education and political affiliation are indeed the most significant factors influencing climate unawareness and denial at the county level, respectively, thus indicating differing predictors for these phenomena. Alternatively, if the study finds that the predictors for both variables are the same, it could suggest that other, less influential factors are at play. Additionally, the study may uncover that the principal predictors vary under certain conditions, such as whether a county is rural or urban.

Such findings would not only enhance our understanding of the factors driving climate unawareness and denial at the county level but also aid in tailoring climate awareness initiatives more effectively. It is crucial, however, to acknowledge that the data from the EDA, derived from Howe et al. (2015), represents estimates rather than direct survey results. Therefore, further research with representative samples from each county would be necessary to internally validate these findings.

This exploratory approach highlights the complexity of public perception towards climate change and the importance of distinguishing between unawareness and outright denial. By dissecting the nuances in predictors based on various county-level characteristics, this study could provide valuable insights into effectively addressing climate change skepticism and fostering a more informed public discourse.

Word Count: ~950

AI Statement: I only used AI in this assignment to help me write comments for my code, paraphrasing to sccunitly write my arguments, and proofread using Grammarly.

[9] #qalmri: I have explained my proposed methodology by keeping in mind the QALMRI approach and all its main points. I have described the logic of my hypothesis and potential results and inferences and how they connect with the method I have used proposed.

References

Ash, E., Boltachka, A., Galletta, S., & Pinna, M. (2023). Media bias and climate change skepticism. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4632854

Howe, P. D., Mildenberger, M., Marlon, J. R., & Leiserowitz, A. (2015c, October 20). *Geographic variation in opinions on climate change at state and local scales in the USA*. Nature Climate Change. https://escholarship.org/uc/item/2bz0416w

Kahan, D. M., Jenkins‐Smith, H., & Braman, D. (2011). Cultural Cognition of Scientific Consensus. *Journal of Risk Research*, *14*(2), 147–174. https://doi.org/10.1080/13669877.2010.511246

Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C.-Y., & Leiserowitz, A. A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, *5*(11), 1014–1020. https://doi.org/10.1038/nclimate2728

McCright, A. M., Marquart-Pyatt, S. T., Shwom, R. L., Brechin, S. R., & Allen, S. (2016). Ideology, capitalism, and climate: Explaining public views about climate change in the United States. *Energy Research & Social Science*, *21*, 180–189. https://doi.org/10.1016/j.erss.2016.08.003

Appendix

R Code:

```
install.packages("maps")

install.packages("mapproj")

install.packages("patchwork")

library(patchwork)

library(maps) #package to create the US map

library(mapproj)

library(tidyverse)

library(ggplot2)

###

data <-
read_csv('https://course-resources.minerva.edu/uploaded_files/mke/Y6azen/howe-2016-data.csv'
)

metadata <-
read_csv('https://course-resources.minerva.edu/uploaded_files/mke/nB7zEY/howe-2016-metadat
a.csv')

### get only the rows in which GeoType are states or counties, and stores them into variables

stateslevel <- filter(data, data$GeoType == 'State')

countieslevel <- filter(data, data$GeoType == 'County')

#splits the GeoName column to seprate State and County Name in the countieslevel dataframe

countieslevel <- countieslevel %>%

  separate(GeoName, into = c("CountyName", "StateName"), sep = ",", extra = "merge")

head(countieslevel)

#makes the state name lower case in stateslevel dataframe

stateslevel$GeoName <- tolower(stateslevel$GeoName)
```

#new columns that stores the difference from national average

stateslevel$Meanhappening <-stateslevel$happening - 70.151

stateslevel$MeanhappeningOppose <-stateslevel$happeningOppose - 12.427

#new column that calculates the percentage of people who do not know that global warming is happening.

stateslevel$DNhappening <- 100 - (stateslevel$happening + stateslevel$happeningOppose)

stateslevel$MeanDNhappening <- (stateslevel$DNhappening - (100 - (70.151 + 12.427)))

#map data for the states

statesmap <- map_data("state")

Renaming 'OldName' to 'NewName'

colnames(statesmap)[colnames(statesmap) == "region"] <- "GeoName"

merge and sort (plots in order, sort ensures states filled in)

stateslevel.geo <- merge(statesmap,stateslevel , sort = FALSE, by = "GeoName")

stateslevel.geo <- stateslevel.geo[order(stateslevel.geo$order),]

p1 <- ggplot(stateslevel.geo, aes(long, lat)) +

geom_polygon(aes(group = group, fill = Meanhappening), color = "black", size = 0.25) +

coord_map() + theme_void() + scale_fill_gradient2(low = "purple", high = "green", mid = "white",

midpoint = 0,

name = "Difference from US average (%)")

plot

p2<-ggplot(stateslevel.geo, aes(long, lat)) +

geom_polygon(aes(group = group, fill = MeanhappeningOppose), color = "black", size = 0.25) +

coord_map() + theme_void() + scale_fill_gradient2(low = "purple", high = "green", mid =

```
"white",
                                   midpoint = 0,
                                   name = "Difference from US average (%)")
# plot
p3 <- ggplot(stateslevel.geo, aes(long, lat)) +
  geom_polygon(aes(group = group, fill = MeanDNhappening), color = "black", size = 0.25) +
  coord_map() + theme_void() + scale_fill_gradient2(low = "purple", high = "green", mid =
"white",
                                   midpoint = 0,
                                   name = "Difference from US average (%)")
p1 <- p1 + theme(
  legend.text = element_text(size = 6), # Smaller legend text
  legend.title = element_text(size = 9), # Smaller legend title
  legend.key.size = unit(0.5, "lines"), # Adjusts the size of the legend keys
  legend.spacing = unit(1, "lines") # Adjusts the spacing between legend keys
)
p2 <- p2 + theme(
  legend.text = element_text(size = 6),
  legend.title = element_text(size = 9),
  legend.key.size = unit(0.5, "lines"),
  legend.spacing = unit(0.5, "lines")
)
p3 <- p3 + theme(
  legend.text = element_text(size = 6),
  legend.title = element_text(size = 9),
```

```
  legend.key.size = unit(0.5, "lines"),
  legend.spacing = unit(0.5, "lines")
)

# Combine plots with unified legend settings
combined_plot <- (p1 + p2 + p3) +
  plot_layout(nrow = 3, guides = "collect") +
  plot_annotation(
    tag_levels = 'A',
    title = "Relative Geographical Variation in Climate Awareness, Uawareness and Denial",
    theme = theme(
      plot.title = element_text(hjust = 0.5, size = 10), # Adjust main title size
      plot.tag = element_text(size = 5) # Adjust subplot label size
    )
  )

# Display the combined plot
combined_plot
##Code for Box Plot
countieslevel$DNhappening <- 100 - (countieslevel$happening +
countieslevel$happeningOppose)
countieslevel$MeanDNhappening <-  (countieslevel$DNhappening - (100 - (70.151 + 12.427)))
# Create a combined dataset with a grouping factor
combined_data <- data.frame(
  values = c(stateslevel$DNhappening, countieslevel$DNhappening),
```

```
  group = factor(rep(c("Counties", "States"), c(length(stateslevel$DNhappening),
length(countieslevel$DNhappening))))

)

# Create the boxplot with different colors

boxplot(values ~ group, data = combined_data, col = c("blue", "red"),

        main = "Variation in % of Climate Unaware People",

        ylab = "% of people",

        names = c("State Level", "County Level"))

#Code for Scatter Plot

ggplot(countieslevel, aes(x = DNhappening, y = happeningOppose)) +

  geom_point(pch = 1) +  # Using open circles for points

  labs(

    x = "% Unaware of Global Warming",

    y = "% Denying Global Warming",

    title = "Global Warming: Awareness Versus Denial at the County Level"

  ) +

  theme_minimal() +  # Using a minimal theme for aesthetics

  theme(

    plot.title = element_text(size = 11)  # Adjust title size

  )

# Calculate the Pearson correlation coefficient

correlation_coefficient <- cor(countieslevel$DNhappening, countieslevel$happeningOppose)

# Square the Pearson correlation coefficient to get R^2

r_squared <- correlation_coefficient^2

# Print the R^2 value
```

```
print(r_squared)
```