Americans on Climate Change

DIGHUM 100 Theory and Methods | Instructor: Dr. Adam Anderson | Vivian Kim | July 1, 2021

A little bit about my motivation...

It's been decades since environmentalists called for changes to preserve our planet and many changes were made in return. Yet the Trump administration, over the course of four years, succeeded in reversing close to *100* rules regarding to air/water pollution, drilling/extraction, just to name a few. My research is not about politics, but rather, I wanted to see if I can identify what kind of factors contribute to have someone stand *against* preserving environment and reasons behind popular and unpopular green activities and policies. The policies mentioned here are not exactly the ones the Trump administration reversed, yet the analysis still gives an idea of the sentiment people have for each policy.

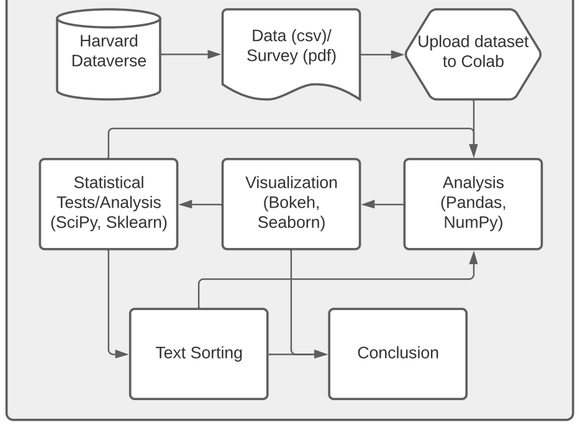
Description/Procedures

How do Americans feel about the ongoing issue of climate change? A survey was given to a sample of people around the world in 2015 asking how much of negative emotions they connect to climate change, what they believe about climate change, how they intend to conserve energy, what kind of morals they have behind climate change, and what they believe should be done in terms of environment-related policies. Along with a short free-response about what they think the purpose of this survey is, their demographics were asked. I focused on respondents from the United States (402 respondents) to answer the following questions:

1) How do gender, age, and political view relate to how they feel about the issue of climate change?
2) What are some of the popular and unpopular green activities and existing/future policies?
3) Are those people who responded "unsure" to the free-response question any different from who didn't?

The dataset is from Harvard Dataverse, and I worked in the Google Colab environment to run my Python codes. In Reference, I have links to my Google Drive and GitHub as well, where I posted the dataset, the actual survey used in this study, and the Jupyter Notebook. For packages, I used NumPy and Pandas for data manipulation, Bokeh and Seaborn for visualization, and SciPy and Scikit-Learn for statistical analysis (2 sample T-test, regression, Kruskal-Wallis, and Dunn's test). For details of the statistical tests and conditions, check my Jupyter Notebook. The roadmap shows the steps I took to create this research project in chronological order. For each question, its statistical analysis follows.

Roadmap



Question 2

A section of the survey asks what kind of actions they intend to take to conserve energy and lists 11 activities: using energy efficient bulbs (**Bulbs**), setting heater to 65° F or below in winter (**Winter**), setting AC to 75° F or above in summer (**Summer**), waiting for a full load before using a washing machine (**Washing**), using a clothesline and not a dryer (**Dryer**), recycling (**Recycling**), buying recycled paper (**Paper**), drinking tap water to reduce plastic (**Water**), using reusable/recycled shopping bags (**Bags**), driving less or not at all (**Drive**), and voting for politicians with environmental policies (**Vote**). I took the average in each category and sorted them from the lowest to highest (1 = *no intention*, 7 = *definitely*). The least popular activities are giving up on using a dryer and driving less/not at all. The most popular activities are waiting to use a washing machine until a full load and recycling. The table shows activities in order from the least popular to most popular.

Activities	Average Score
Dryer	3.353234
Drive	3.850746
Winter	4.427861
Summer	4.664179
Paper	4.796020
Vote	5.014963
Bags	5.019900
Water	5.268657
Bulbs	5.485075
Recycling	5.743781
Washing	5.791045

Question 1

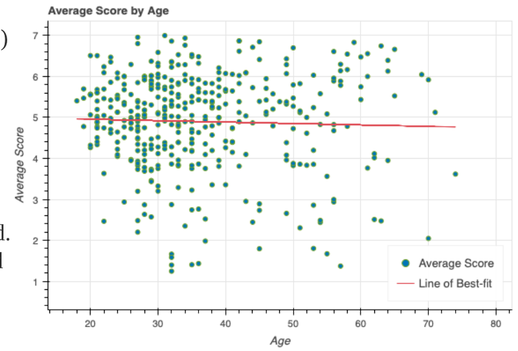
I grouped the dataset by gender, age, and political standpoint and got the average of the sentiment scores, which are in a scale of *1* to *7* (1 = *apathetic*, 7 = *mindful* of climate change). I conducted statistical tests to further explore the relationships of given variables.

Gender

From the graph, we can see that the average scores for women and men are about 4.980 and 4.801 respectively, which doesn't seem too far off from each other at a glance. Women in this sample answered slightly more empathetic about environmental issues. To see if their difference hold any statistical meanings, I conducted a 2 sample T-test. I first checked if conditions were met (see Jupyter Notebook for more), I proceeded to use `ttest_ind` in the `stats` function of SciPy. The null hypothesis is that the two means are equal whereas the alternative hypothesis would be that their difference is statistically significant, or they are not equal. The resulting P-value is 0.1315, which at 5% level, I **failed** to reject the null hypothesis. Their difference in means is not significant and gives evidence that the two means are equal.

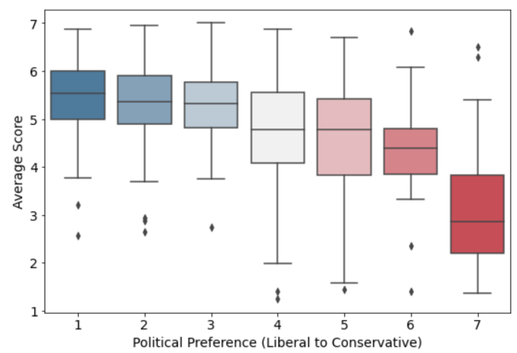
Age

The graph shows individuals' scores based on their age. We can see that data points in the upper left corner side are more clustered than the rest, meaning that younger people (around the age 20 to 40) scored higher than the older people. Interesting to note that the data points are more widely spread in older age groups, yet still more data points in the upper half than the lower half (half-point at 4). The regression line is also obtained using `linregress` in `stats` function of SciPy then plotted: `y = -0.0035x + 5.026`. The correlation is -0.0346, meaning the relationship between age and the average score is very weak and they are almost uncorrelated. To see how much of data this regression line represents, I computed the R-square value, 0.0012. Less than 1% of the data is explained by the regression line! This is due to how widely scattered the data points are and their linear relationship is **uncorrelated**.



Political Preference

Similar to the gender graph, but this time I used a bar plot to show the seven political groups from 1 to 7 (1 = *strongly liberal*, 7 = *strongly conservative*). Each box represents a group in order, 1 to 7. The boxes are also shown in different colors: blue and red for democratic and republican parties respectively. Due to the different number of sample sizes and wide variation of standard deviations, I used a non-parametric version (compares *medians* instead of means) of the one-way ANOVA (Analysis of Variance), called *Kruskal-Wallis* test. For such a task, I used `stats.kruskal` from SciPy. The obtained P-value is 8.344e-13, which is much smaller than 0.05, at 5% level, which means not all medians in the group are equal and their difference is not due to a random chance. I went ahead and conducted another test, the *Dunn's test* in `sklearn.posthoc_dunn` to compare adjusted P-values pairwise. Out of 11 combination of groups with statistically significant differences, the top three are: 1 and 7 (5.404e-8), 2 and 7 (1.186e-7), and 3 and 7 (8.284e-6). This shows the differences in medians of *any liberal* groups and *strongly conservative* group are highly statistically **significant**.

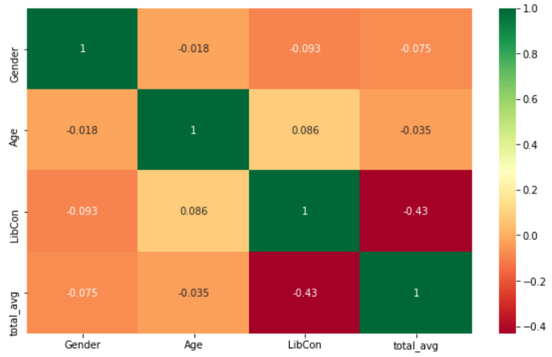


Question 1 (continued)

After I analyzed the three variables (gender, age, political preference) in the previous page, I wanted to compare those three side-by-side this time as I found some of the results not quite satisfying. How much does each feature contribute to a respondent's average sentiment score?

Feature Comparison

I compared the three features (gender, age, and political preference) using the built-in function `corr` of Pandas and the `heatmap` from Seaborn for visualization. The `corr` evaluates pairwise correlations of columns and `heatmap` gives each combination block a color, depending on the correlation value, defined in the color bar on the right. This way, I can easily visualize and compare the correlations of the features. They all are either *negatively* correlated which is what we expected from the previous page. In exact values, correlations of `total_avg` follows: with gender (`Gender`), -0.075; with age (`Age`), -0.035; with political preference (`LibCon`), -0.430. Although a correlation value of -0.430, for political view, is considered moderately correlated, notice how this correlation is multiples larger than the other two correlations. Gender has a slightly larger correlation, however, both gender and age have very small, closer to 0 correlations, or uncorrelated.



Question 4

Upon manually checking the responses of the free-response question asking if they can guess what the purpose of the survey is, I figured most people guessed something along the lines of environment. Some respondents guessed it was a psychological research to see how they answer to the questions after reading the first intro page (which it was, actually). However, about 3.98% of the people responded "unsure" to this question. There are 16 respondents in this group and I compared their mean score (4.850 for "unsure") to the rest (4.903 for "sure") who responded not "unsure." Though I believe some responded this way out of pure laziness, I thought it would still worth to find out if the difference in their scores has a significant meaning statistically. I, again, conducted a *2 sample T-test*. It was a bit trickier than Question 1, since the "unsure" group is much smaller than the "sure" group. However, the sample size of 16 is considered large enough, I proceeded to use `stats.ttest_ind` of `SciPy`. The P-value is evaluated to be 0.8628, which, at 5% level, I fail to reject the null hypothesis of their means being equal. The table on Question 4 has the list of the people and their demographics in the "unsure" group and their actual responses to the question.

Gender	age	LibCon	Purpose
male	31	3	I don't know
male	21	1	I can't even begin to guess.
female	19	2	no idea
male	24	3	Not sure
female	21	1	I'm not sure
male	34	1	I am not sure.
male	23	4	unsure
male	43	4	I'm not sure.
female	34	2	unsure
female	24	4	not sure
male	23	1	not sure really
male	32	1	i have no idea
male	29	1	I AM UNSURE
female	27	6	\I am not sure"
female	29	4	I really dont know
female	28	3	i'm not sure

Interpretation

1) After conducing 2 sample T-tests on two gender groups, it's found that gender does *not* play a significant role in whether or not the person cares and actively participates in cherishing our environment. Similarly, the age and average score are analyzed to be *uncorrelated*. I would have been disappointed to find no linkage among my chosen features, yet I was able to find a relationship between the political preference and average score. I observed that the *more* conservative someone is, he/she is *more* likely to be indifferent to protecting environment. After two statistical tests, the median differences in liberal and conservative groups were **highly statistically significant** as well. In particular, the median differences of *any* liberal groups to the strongly conservative group are all highly significant. The correlations of features showed political preference and the average score has its correlation much *higher* than the rest.
2) I analyzed the most popular and least popular intended environmental activities. From the most to least in order, the list goes like this (see previous page for which activity each word represents): washing, recycling, bulbs, water, bags, vote, paper, summer, winter, drive, and dryer. Perhaps geographically and culturally it's harder for Americans to give up on driving due various reasons like safety or underdeveloped public transportation, yet other activities such as giving up on using a dryer or setting an A/C or heater to a certain temperature can be easily achieved.
3) The most supported to least supported policies follows (refer to Question 3 for which policy each word represents): reuse, hybrid, cup, carpool, plastic, MPG, and commute. It's clear that people support *rewards* more than *punishments*.
4) In my opinion, the people who could not guess can be a *bigger* threat than the ones guessed something because it may have mean that the attempt to educate people on this matter was *insufficient*. However, when I conducted a 2 sample T-test to compare the difference in their average scores ("unsure" group and "sure" group), I saw that their difference is statistically **insignificant**. I planned to analyze further about this "unsure" group, but since the result of the T-test showed that they are no different, I stopped here.

Conclusion

The fundamental question I wanted to find an answer was, "what is different about an environmentally-unfriendly person?" It was not the generation gap or gender difference, but *politics*. Though I believe one's political preference is deeply related to his or her level of education, income, region, etc., which can be the next research topic. In particular, perhaps this is due to lack of education that some people don't believe in environmental issues and is caused by a low income. Many factors are also intertwined to form a person's political viewpoint, yet our society has witnessed politicians threatening our environment to gain political support. In order to preserve environment, this, as a society, must raise moral concerns. From the result regarding to green activities, how these activities get promoted and exposed to the public is the key, especially when there is an alternative method already available. For example, it might not be so clear that opting out of using a dryer cuts over a ton of greenhouse gas and synthetic fibers per year, whereas we most likely can connect recycling as an environmentally-friendly activity from how recycling has been advertised. As for policies, it's obvious that people want to get paid instead of getting fined. Using this logic of rewarding those who participate, green policies can be enforced more effectively by motivating people and not only relying on volunteers. On a brighter note, people *can* change their beliefs, opinions, and political views. It might take some time, yet as a community we must continue to raise awareness about environmental issues and reward those who engage in such activities while *penalizing* those purposely harming our planet.

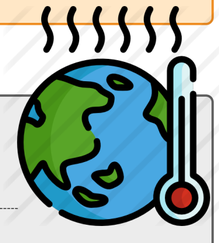
Further Questions

1) How can we convince more people to participate in ensuring environmental sustainability?
2) What can we do to promote green activities more effectively?
3) Can policies be used as a tool to motivate people to do more environmentally-friendly activities?
4) How can we restrain politicians from threatening out environment for political reasons?
5) What other factors contribute to a person's environmental-unfriendliness? (ie. level of education, income, etc.)

Question 3

Part 7 of the survey is about whether or not respondents agreed or disagreed to a certain policy. These questions are answered in *Yes* or *No* (1 = *Yes*, 2 = *No*). The policies mentioned follows: a fee for a disposable plastic shopping bag (**Plastic**), a rebate for a reusable shopping bag (**Reuse**), a rebate for a reusable cup (**Cup**), a commuter tax for driving a car alone during rush hour (**Commute**), tax credits for hybrid/electric cars (**Hybrid**), a fee for owning a car with a low fuel efficiency (**MPG**), and extra paid time off for carpooling employees (**Carpool**). I computed the average value for each policy and the table on the right shows the least popular to the most popular in order. The popular policies are a rebate for a reusable shopping bag and tax credits on hybrid vehicles, and the unpopular ones are a fee for fuel-inefficient cars and a commuter tax for people driving during the rush hour.

Policies	Average Score
Reuse	1.136816
Hybrid	1.166667
Cup	1.194030
Carpool	1.417910
Plastic	1.509950
MPG	1.554726
Commute	1.788557



References

Data: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8Z7M4G
Google Drive: https://drive.google.com/drive/folders/1S5Ys2fv0L0xWvG6Gx8jBeZYuqU6_a8sj?usp=sharing
Notebook: https://colab.research.google.com/drive/1Y31yjvjlNxSN529JSBmEVg1eJHyNXlNE?usp=sharing
GitHub: https://github.com/vivianskim20/climate-change

Images: https://www.nursingcenter.com/ncblog/november-2019/climate-change
https://www.clipartmax.com/middle/m2H7i8G6b1G6b1i8_global-warming-global-warming-clipart/
https://www.clipartmax.com/middle/m2i8d3A0H7Z5Z5N4_global-warming-global-warming/
https://www.clipartmax.com/middle/m2i8G6N4H7m2A0A0_tree-snow-winter-ice-cold-forest-icon-snow-forest-icon/
https://unsplash.com/photos/UwBrS-qRMHo
Articles: https://www.nursingcenter.com/ncblog/november-2019/climate-change
https://www.nytimes.com/interactive/2020/climate/trump-environment-rollbacks-list.html
https://cleanhomeguide.co.uk/environmental-impact-of-laundry/ https://particle.scitech.org.au/people/impact-of-air-conditioners-on-the-environment/
Technical: https://www.schemecolor.com/color/monochromatic
https://seaborn.pydata.org/generated/seaborn.diverging_palette.html
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html,
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html#scipy.stats.kruskal
https://www.statology.org/anova-assumptions/
https://support.minitab.com/en-us/minitab/19/help-and-how-to/statistics/nonparametrics/how-to/kruskal-wallis-test/interpret-the-results/all-statistics/
https://www.statology.org/two-sample-t-test-python/
https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e
https://stackoverflow.com/questions/51198184/seaborn-heatmap-not-showing-columns-converted-from-string-to-numerical
https://www.geeksforgeeks.org/python-pandas-dataframe-corr/
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.corr.html
https://www.statology.org/dunns-test-python/
https://scikit-posthocs.readthedocs.io/en/latest/generated/scikit_posthocs.posthoc_dunn/
https://www.statology.org/family-wise-error-rate/
Riddell, Corinne. "Comparing Two Means." "Comparison of Many Means"
Public Health 142. 2020. University of California, Berkeley.