

Final Project Report



Global Warming under Big Data Analysis

Team #Frau mit Sonnenschirm:

Pengzi Li

Xuanzhao Li

Junan Mao(Sabrina)

Muhan Wang

Shumei Chen

Special Topics: Managing Big Data / MASY1-GC 5000-101

May 6, 2019

Abstract

Climate change has been recognized as the most prominent environmental problem of the 21st century. Before educating and letting the general public be aware of the importance of this issue, we need to know their attitude toward the global warming problem first.

The objection of this project is to identify how the public responds emotionally to global warming. The hypothesis of this study is that people will have higher perceptibility of global warming and realize the occurring of global warming faster when they experienced one of the factors.

Data is collected from the year 1998. To estimate how many people believe the existence of global warming, google ngram viewer data is used. The carbon dioxide data, the global climate data are between 1998 and 2008, they are collected from the US Environmental Protection Agency And global historical climate network. The sentimental data is collected on twitter using key words “global warming”. Linear regression and Naïve Bayes supervised classification are used to analyze the project data. The study concludes with some brief prospects that people believe in the existence of global warming.

Key words: global warming, Twitter, Google Ngram, sentiment analysis, linear regression, Naïve Bayes supervised classification

Content

Abstract

1. Introduction

2. Problem Statement

3. Data

4. Method

4.1 Linear regression

4.2 Natural Language - Naive Bayes Supervised Classification

5. Conclusion

5.1 Major findings

5.1.1 Exploratory Data Analysis

5.1.2 Linear Regression

5.1.3 Sentiment Analysis

5.2 Limitations

Appendix

References

1. Introduction

Global warming is a long-term rise in the average temperature of the Earth's climate system, an aspect of climate change shown by temperature measurements and by multiple effects of the warming. Is global warming really happening? The answer is a definite yes. Two major reports released since 2014, the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) and the Climate Science Special Report (2017) have stated unequivocally that temperatures are rising.

Since 1979, the rate of warming has approximately doubled. Since 1950, the number of cold days and nights has decreased, and the number of warm days and nights has increased (Hartmann et al., 2013). The global-mean surface temperature has risen by about 0.7°C since the second half of the nineteenth century, which is mainly caused by increased greenhouse gases (Meehl & Coauthors, 2012). Greenhouse gases (GHGs) are trace atmospheric gases that trap heat on Earth, preventing it from escaping into space (Pan, 2019). The primary greenhouse gases are water vapor, which causes about 36–70% of the greenhouse effect; carbon dioxide (carbon dioxide), which causes 9–26%; methane (CH₄), which causes 4–9%; and ozone (O₃), which causes 3–7%.

Glaciers are melting, sea levels are rising, cloud forests are dying, and wildlife is scrambling to keep pace. The result is often called global warming. It is causing a set of changes to the Earth's climate while the public thinks of global warming and climate change as synonyms, scientists prefer to use “climate change” when describing the complex shifts affecting earth's weather and climate systems. Some areas actually can be cold even as the planet warms globally.

The use of “global warming” and “climate change” has become very visible to the public that they respond to all of the news and publicity about climate, expressing support or disbelief. Some individuals experienced cold weather keep skeptical about Global Warming is happening.

Back on the evening of December 28, 2017, someone wrote on his infamous Twitter feed: "In the East, it could be the COLDEST New Year's Eve on record. Perhaps we could use a little bit of that good old Global Warming that our Country, but not other countries, was going to pay TRILLIONS OF DOLLARS to protect against."

On social media like Twitter, the large amount of data available for mining to formulate a response to public opinion regarding the subject being assessed. In modern society understanding how the public responds regarding complex issues of societal importance. Twitter data is indicative of how people refer to topics of interest, in a manner that is very linguistically restricted. As well, Twitter used as a platform for verbal expression of emotional responses. Due to the restrictions on tweet size (each tweet can only be 140 characters in length), it is necessary to be more direct in dealing with topics of interest to the tweeter. Therefore, the tweets are linguistically more emotionally charged and can be used to define a level of emotional response by the tweeter.

At the same time, Google Ngram Viewer is used to reflect the popular opinions. The Google Ngram Viewer is an online search engine that charts the frequencies of any set of comma-delimited search strings using a yearly count of n-grams found in sources printed between 1500 and 2008 in Google's text corpora. It is one method of measuring the popularity of a subject within the population. Keywords in printed materials are topic sensitive, and therefore indicate the level of public knowledge regarding the searched topic.

The phrase “global warming” here is a unilateral term that indicates a level of awareness about the issue which is indicative of the public’s interest and attitude in that subject.

In previous research, Lineman (2015) used public databases of Google queries, Google Trend, and Twitter to combine search trend and sentiment analysis. They identified changes in public perception and interest. Moore (2019) studied the relationship between remarkability of temperature anomalies and public opinion of climate change. The research discussed the human evaluation of weather as either normal or abnormal will also be influenced factors including expectations, memory limitations, and cognitive biases. The remarkability of particular temperatures changes rapidly with repeated exposure.

The following study assessed relative printed materials and new media for global warming and climate change to determine knowledge and awareness of these terms. And then further study the shift in public recognition at different times or detect what elements had influenced some individuals' doubt on global warming.

2. Problem Statement

As we introduced above, global warming is a significant concern nowadays, and weather conditions are changing. In fact, since the beginning of the industrial revolution, the burning of fossil fuels has increased the non-natural emissions of carbon dioxide to the atmosphere (Lindsey, 2018). Carbon dioxide is a greenhouse gas that absorbs the infrared radiation produced by the reflection of the sunlight on the Earth's surface, trapping the heat in the atmosphere. Global warming and the associated global climate changes are actively being used as the subject of intensive research due to their critical impact on the social, economic, ecosystem, agriculture and health aspects of human life.

There is the global warming trend observed by scientist since the mid-20th century to the human expansion of the "greenhouse effect" which also called warming that results when the atmosphere traps heat radiating from Earth toward space(Naomi, 2004). There are also certain gases in the atmosphere block heat from escaping, and some of them remain semi-permanently in the atmosphere and do not respond to changes in temperature which leads to climate change eventually, like water vapor and carbon dioxide. According to the Environmental Protection Agency, carbon dioxide concentrations in the atmosphere have risen from 280 parts per million in the 18th century to 390 parts per million in 2010. Human activities release more than 30 billion tons of carbon dioxide each year, or 135 times as much as volcanoes. Except for production and emission of carbon dioxide, and constant globally weather change, we also screen out another main factor of global warming for our study — global sea level. There is some factual evidence for rapid “global warming.” According to data from NASA's Gravity Recovery and Climate Experiment, it indicates that Greenland lost an average of 286 billion tons of ice per year between 1993 and 2016, while Antarctica lost about 127 billion tons of ice per year during the same period. The rate of Antarctica ice mass loss has tripled in the last decade(Ramp-Up in Antarctica Ice Lose Speeds Sea Level Rise, 2018). There are around 8 inches rose of global sea level in the previous century. The sea level rate in the last two decades is surprisingly nearly double the scale of the last century, and it is accelerating slightly every year (R. S. Nerem, 2018).

Due to all the facts related to our main concern — the factor of global warming, the main problem of this study is to estimate which global warming factors would most enhance people's perceptibility of global warming and also influence people's beliefs about Is global warming is still happening. In this study, we are going to utilize Google Ngram Viewer and sentimental analysis on Twitter to obtain data from 1998 to 2008 and focus on analyzing these factors which have high effects on global warming: global climate(weather), real-world carbon dioxide emission and global sea level. Moreover, for observing the effects of global climate change during that time, we controlled variable and selected maximum temperature, minimum temperature and snow depth per day as main factors to analyze as well. So, this study hypothesizes that people will have higher perceptibility of global warming and realize the occurring of global warming faster when they experienced one of the factors include: global climate change, announced the high-level carbon dioxide emission and production, and the rise of global sea level.

3. Data

In the first part of our project, we wanted to estimate what global warming factors would influence people's beliefs in global warming. Beginning in 1998, people started to suspect the existence of global warming based on the satellite record estimates of the temperature of the atmosphere, which did not show a warming trend at the time. But in fact, the warming trend continued (Lindsey, 2018). So in this part of the project, our data will start in 1998.

One way to estimate how many people believe the existence of global warming is by using Google Ngram Viewer data. The key word that we use to obtain data from n-gram is “global warming exists” and “global warming doesn't exist.” The limitation of this data source is that it only includes data until 2008; how people think about the existence of global warming after 2008 cannot be detected. Hence, in this part of our project, our date range is between 1998 and 2008.

Carbon dioxide is the main cause of global warming, which is mainly emitted through deforestation and burning fossil fuels(Change, 2019). We employ global carbon dioxide production and emission from gaseous fuel consumption between 1998 to 2008 from the Carbon Dioxide Information Analysis Center. The features we choose are the production of carbon dioxide (in thousands) and real-world carbon dioxide emission (g/mi).

Two primary factors that prove the existence of global warming are global temperature and sea level (Climate Change Evidence: How Do We Know? (n.d.), 2019). The sea level data we use in this project is the global average sea level change between 1998 to 2008 from the US Environmental Protection Agency using data from the Commonwealth Scientific and Industrial Research Organization (CSIRO). In this dataset, we choose CSIRO adjusted sea level data, which shows cumulative changes (in inches) in sea level for the world's oceans based on a combination of long-term tide gauge measurements and recent satellite measurements.

For the global climate data, we use the global historical climate network including daily land surface observations from around the world. Low temperature and snow depth might lead people to suspect the existence of global warming (Whatever happened to global warming? How freezing temperatures are starting to shatter climate change theory. (n.d.), 2019), so we only choose maximum temperature, minimum temperature and snow depth per day, and then aggregate them into year levels, from 1998 to 2008.

The second part of our project is sentiment analysis on Twitter through Earth Day (April 22, 2019). The keyword we use for downloading twitter data through API keys is “global warming.” The training set we obtained is from Figure Eight, created by Kent Cavender-Bares. Cavender evaluated tweets for the belief in the existence of global warming or climate change. The possible answers were “Yes” if the tweet suggests global warming is occurring, “No” if the tweet suggests global warming is not occurring, and “I can’t tell” if the tweet is ambiguous or unrelated to global warming.

4. Method

4.1 Linear regression

In the first part of our project, we construct a linear regression model to test the causality of the belief in global warming. The dependent variable is the proportion of people believing in the existence of global warming between 1998 to 2008; the independent variables are the weather features, carbon dioxide features, and sea level feature.

4.2 Natural Language - Naive Bayes Supervised Classification

After downloading the tweets which contain “global warming” on earth day, the first thing to do is split them into words using a stopwords text file and then remove duplicate words. By deploying naive bayes classifiers from the natural language toolkit, we then have the dataset that contains the date and time, the content of each tweet, and sentiment result which evaluated tweets for belief in the existence of global warming. Since we only care about the number of people believing in the existence of global warming or not, we remove the N/A value from the sentiment features.

5. Conclusion

5.1 Major Findings

This project is an analysis of people believe in the existence of global warming, and the major findings are combined the n-gram linear regression model and Twitter sentiment analysis findings.

5.1.1 Exploratory Data Analysis

On the analysis of n-gram data, the project analyzes 6 main factors had affected people’s judgment on the existence of global warming, which are the maximum temperature (Tmax), minimum temperature (Tmin), snow depth, real-world fuel economy emission, cumulative sea

level, and real-world carbon dioxide emission (carbon dioxide). The project attempts to explore the factor which is a positive or negative correlation to the collected n-gram data. There are five positive correlations to the proportion of printed sources indicate global warming exists: Tmax, Tmin, snow depth, real-world fuel economy emission, and cumulative sea level. And only one negative correlation is real-world carbon dioxide emission. On the contrary, the proportion of printed sources indicate global warming does not exist opposite results that its positive correlation is real-world carbon dioxide emission; its negative correlations are Tmax, Tmin, snow depth, real-world fuel economy emission, and cumulative sea level (see Graph 1).

Moreover, in this study, the team also addresses that the proportion of carbon dioxide produced each year has the same trend as the proportion of printed sources, which highly indicate that global warming exists (see Line graph 1). And the trend of global carbon dioxide emission per year has a lag effect on n-gram data (see Line graph 2). However, there no notable relation between average maximum temperatures per year, average minimum temperatures per year, average snow depth per year, sea level per year, and global fuel economy emission to n-gram features (see Line graph 3,4,5,6,7).

5.1.2 Linear Regression

From the correlation matrix (See Graph 1), we noticed that the correlation between Tmax and Tmin, carbon dioxide production and carbon dioxide emission, fuel emission are highly correlated. In case of overfitting problem, we only choose maximum temperature (Tmax), carbon dioxide production, and sea level as our independent variables. We use the proportion of printed source that has a positive attitude to the existence of global warming as our dependent variable.

The R-square of the linear regression model is 0.595, which indicates that the independent variables can explain 59.5% of the dependent variable. We set a significant level at 0.05, so the only feature that has a substantial effect on the dependent variable is carbon dioxide production, which has p-value equal to 0.01. Therefore, we can conclude that people are sensitive about the existence of global warming when the carbon dioxide production change.

5.1.3 Sentiment Analysis

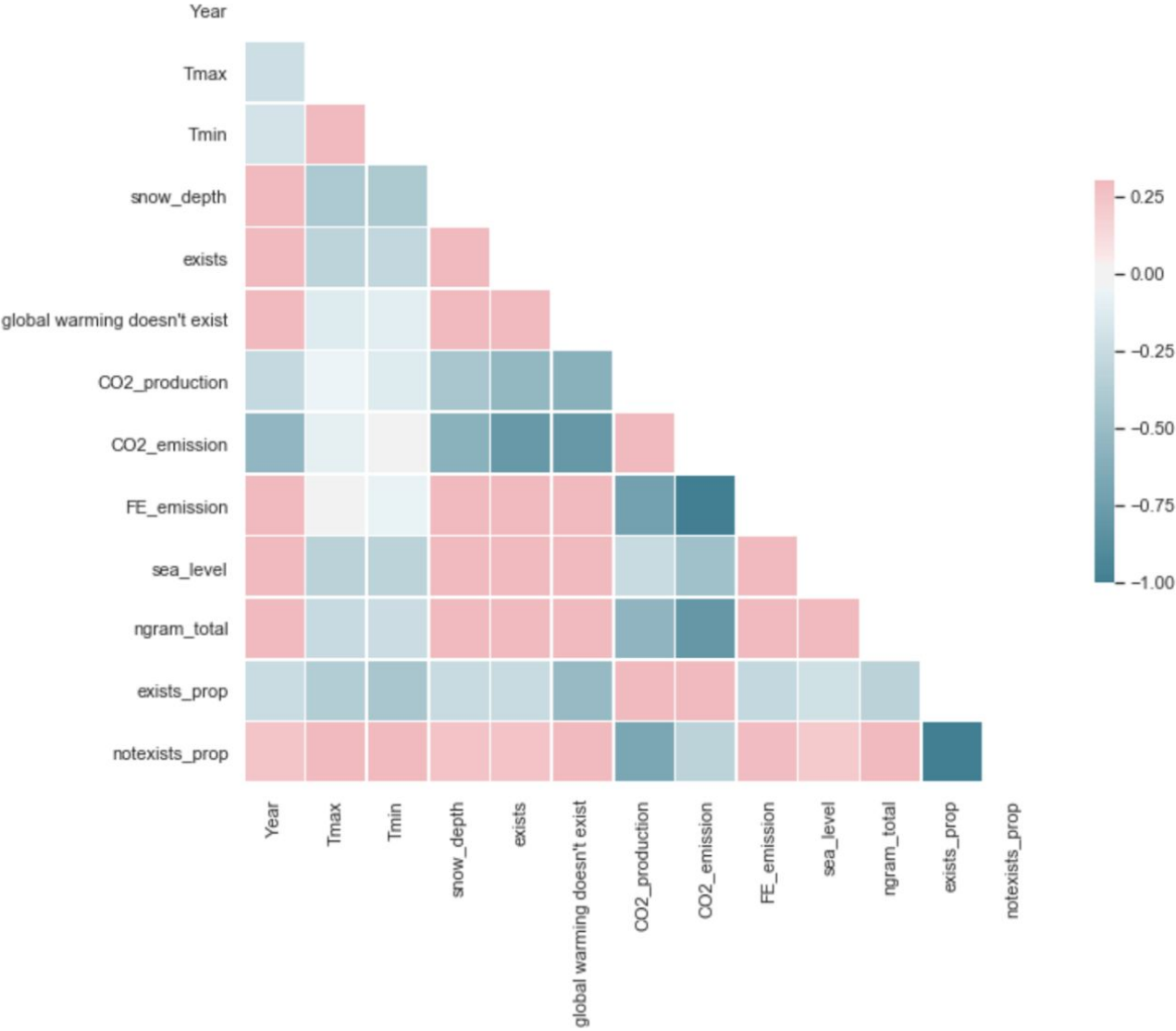
On the analysis of Twitter sentiment analysis of people believe global warming exists or not on Earth Day in 2019, most of the people has the positive attitude towards the existence of global warming (see Figure 2). The team defines the Earth Day times into hours, minutes and seconds for analysis. We explored out that people posted tweets that indicate they believe global warming exists or not exists on Earth day April 22, 2019, are fluctuated between 0 a.m. and 3 a.m. After 3 a.m. both features are rapidly increasing, but most of the people believe that global warming is happening. The number of tweets that believe global warming is exists reached maximum at 8 p.m, but keep decreasing after that (see Line graph 8).

5.2 Limitations

There are some limitations to the project. First of all, this research data is built on small comparative databases, especially the n-gram only from 1998 to 2008 which is not having a high universality to all effect on global warming. Secondly, all the n-gram and Earth Day tweets are searched in the English version, which will also bring the limitation of not general enough since there are many other languages of these data. Thirdly, the correlations between n-gram and factors are not strong, and it's very light correlations, but the linear regression still have multicollinear problem. Last but not least, Google and Twitter have climate skeptic censorship which has suppressed or banned the content of denying climate change, and it will affect the accuracy of the analysis result.

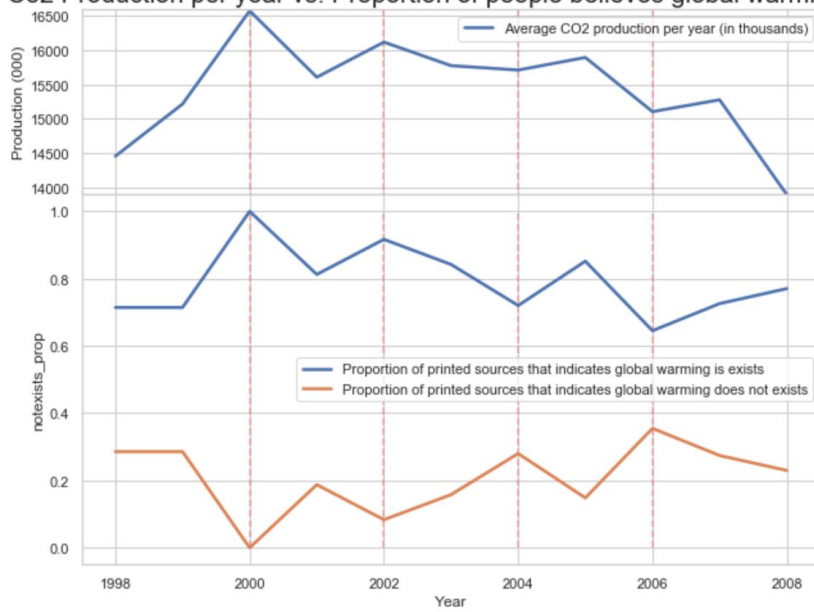
Appendix

Graph 1.



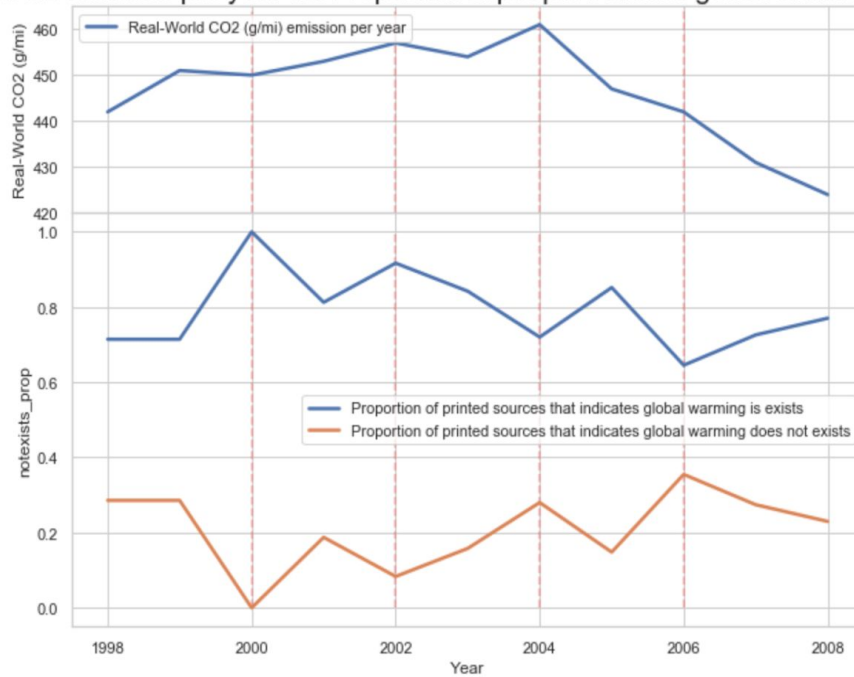
Line graph 1.

Global Co2 Production per year vs. Proportion of people believes global warming exists or not



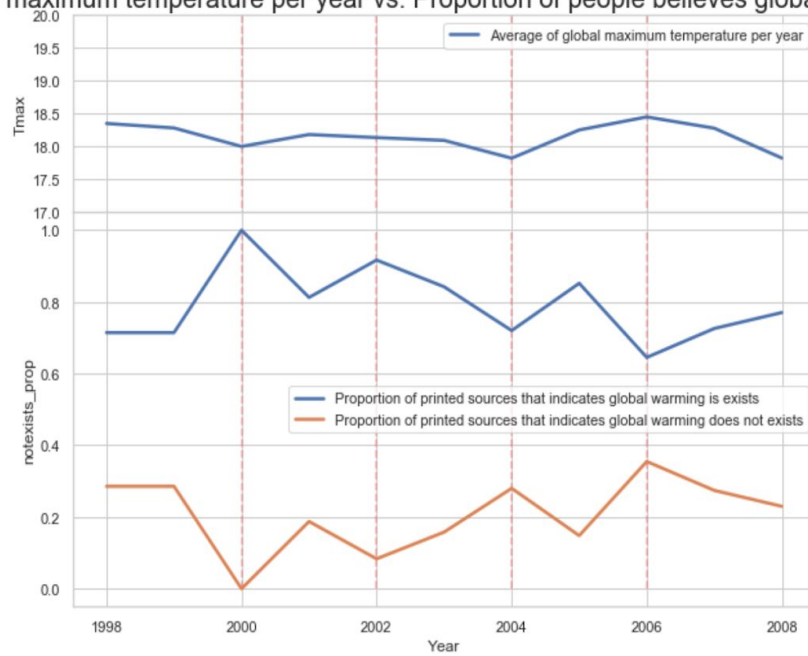
Line graph 2.

Global CO2 emission per year vs. Proportion of people believes global warming exists or not



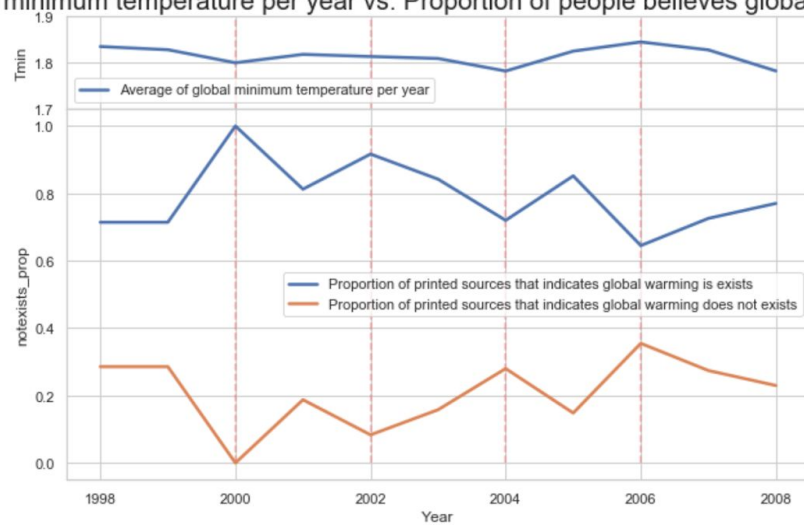
Line graph 3.

Average global maximum temperature per year vs. Proportion of people believes global warming exists or not



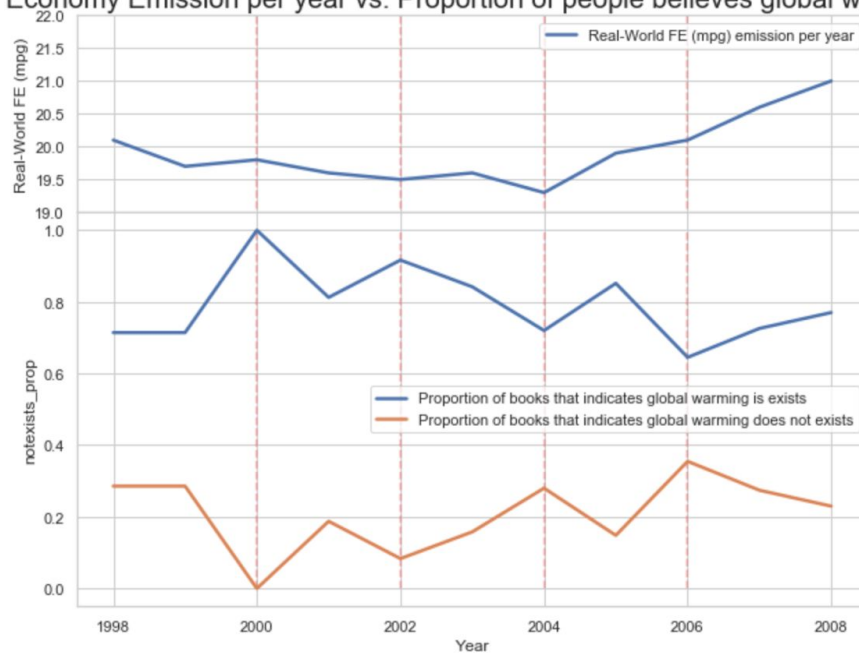
Line graph 4.

Average global minimum temperature per year vs. Proportion of people believes global warming exists or not



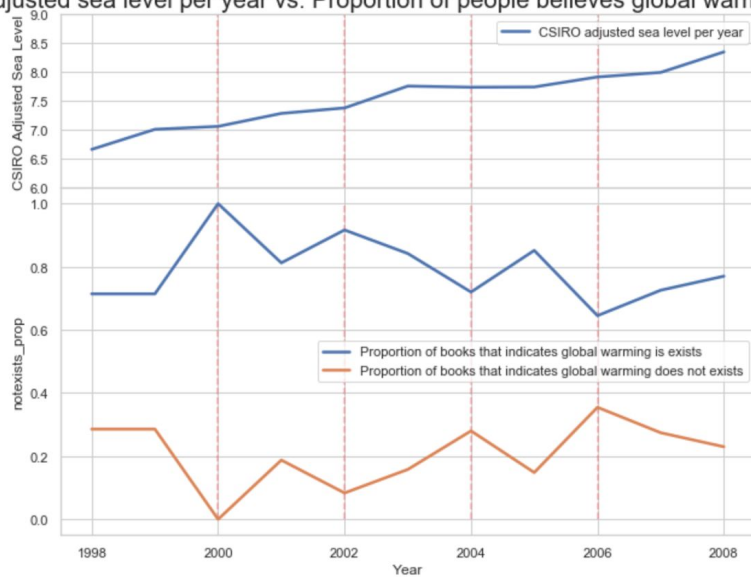
Line graph 5.

Global Fuel Economy Emission per year vs. Proportion of people believes global warming exists or not



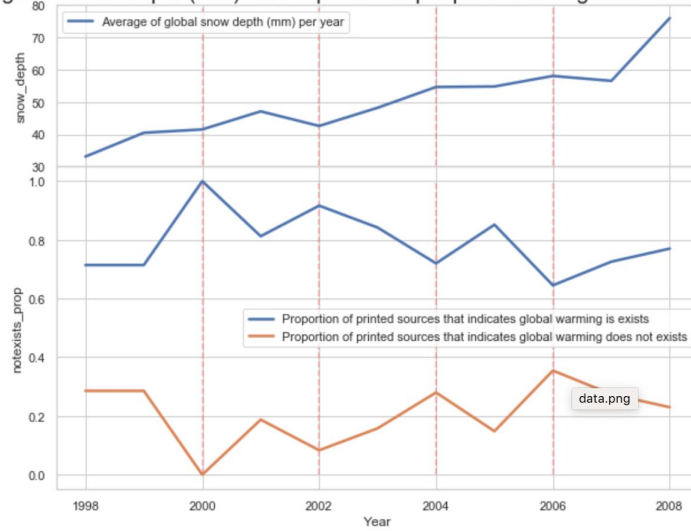
Line graph 6.

CSIRO adjusted sea level per year vs. Proportion of people believes global warming exists or not



Line graph 7.

Average global snow depth (mm) vs. Proportion of people believes global warming exists or not



Line graph 8.

Sentiment Analysis of people believe global warming exists or not on Earth day (Apr 22, 2019)

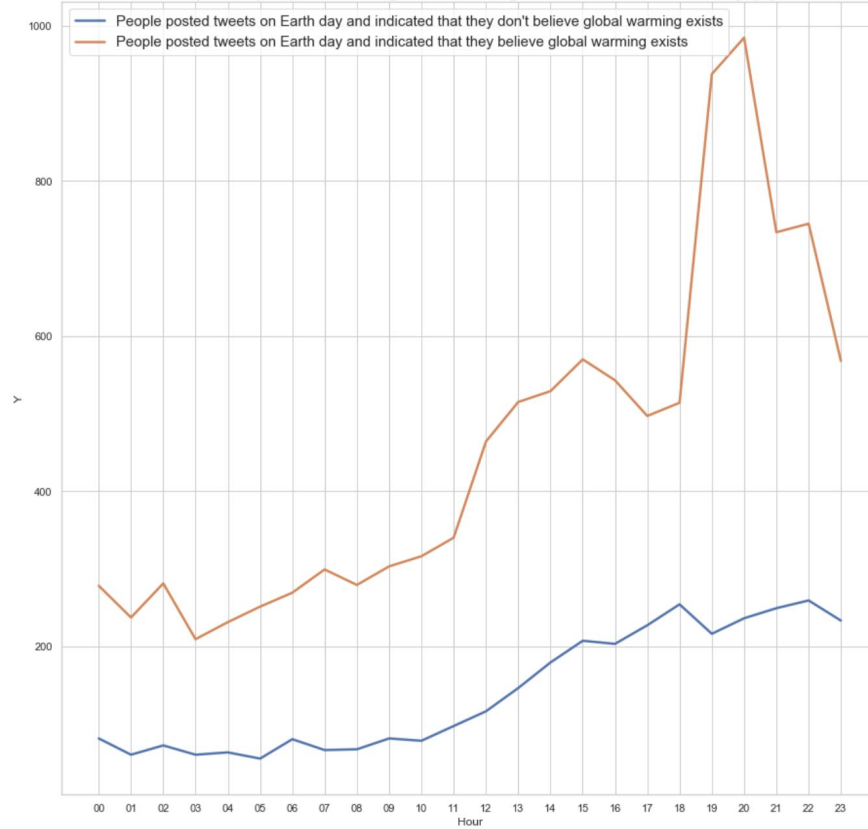
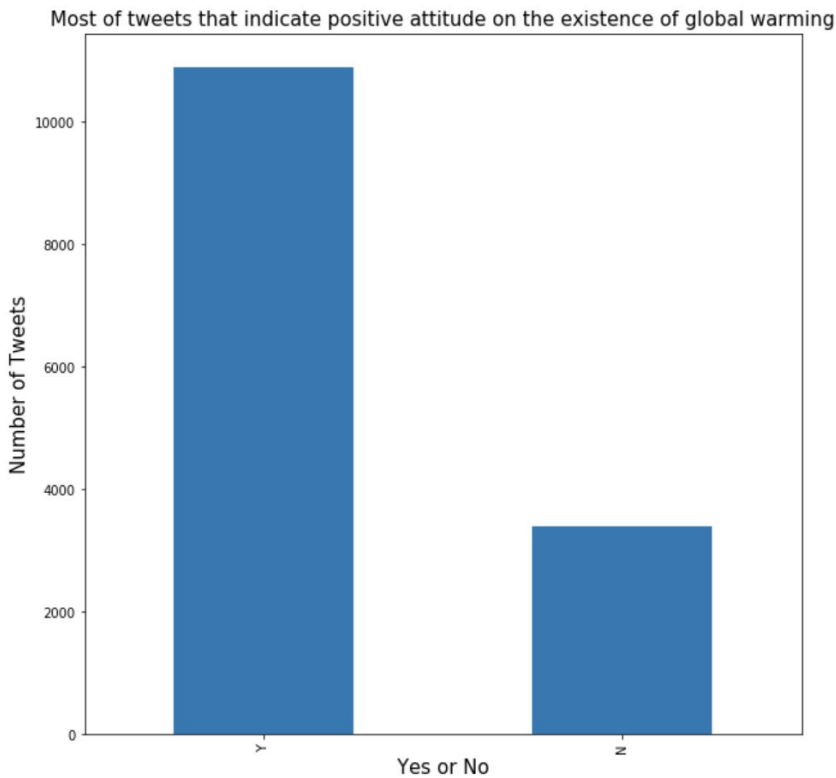


Figure 1. Linear Regression Model

OLS Regression Results						
Dep. Variable:	exists_prop	R-squared:	0.773			
Model:	OLS	Adj. R-squared:	0.621			
Method:	Least Squares	F-statistic:	5.095			
Date:	Mon, 06 May 2019	Prob (F-statistic):	0.0390			
Time:	17:56:31	Log-Likelihood:	17.915			
No. Observations:	11	AIC:	-25.83			
Df Residuals:	6	BIC:	-23.84			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.7388	2.513	0.294	0.779	-5.410	6.888
Tmax	-0.2393	0.108	-2.209	0.069	-0.504	0.026
CO2_production	0.0001	3.93e-05	3.673	0.010	4.82e-05	0.000
FE_emission	0.1434	0.066	2.164	0.074	-0.019	0.306
sea_level	-0.0910	0.051	-1.782	0.125	-0.216	0.034
Omnibus:	0.901	Durbin-Watson:	2.301			
Prob(Omnibus):	0.637	Jarque-Bera (JB):	0.728			
Skew:	-0.351	Prob(JB):	0.695			
Kurtosis:	1.954	Cond. No.	2.00e+06			

Figure 2. Frequency bar plot



References

Change, N. G. C. (2019). Carbon Dioxide Concentration | NASA Global Climate Change.

Retrieved from <https://climate.nasa.gov/vital-signs/carbon-dioxide>

Christina Nunez, (2019). *Causes and Effects of Climate Change*. National Geographic.

Retrieved from

<https://www.nationalgeographic.com/environment/global-warming/global-warming-overview/>

Climate Change: Vital Signs of the Planet. (2019). *Climate Change Evidence: How Do We*

Know? Retrieved from <https://climate.nasa.gov/evidence>

Hartmann, D. L.; Klein Tank, A. M. G.; Rusticucci, M.; Alexander, L. V.; Brönnimann, S.; Charabi, Y.; Dentener, F. J.; Dlugokencky, E. J.; Easterling, D. R.; Kaplan, A.; Soden, B. J.; Thorne, P. W.; Wild, M.; Zhai, P. M. (2013), Chapter 2: Observations: Atmosphere and Surface (*PDF*) in IPCC AR5 WG1 2013.

IPCC. (2014). *Fifth Assessment Report*.

Leo Goldstein. (2018). *Censorship by Google, Twitter, and Microsoft LinkedIn*. Retrieved from <https://wattsupwiththat.com/2018/01/01/climate-skeptic-censorship-by-google-twitter-and-microsoft-linkedin/>

Lineman, M., Do, Y., Kim, J. Y., & Joo, G.-J. (2015). *Talking about Climate Change and Global Warming*. PLoS ONE, 10(9), 1–12. Retrieved from <https://doi-org.proxy.library.nyu.edu/10.1371/journal.pone.0138996>

Lindsey, Rebecca. (2018). In *Did global warming stop in 1998?*. Retrieved from <https://www.climate.gov/news-features/climate-qa/did-global-warming-stop-1998>

Lindsey, Rebecca. (2018). In *Climate Change: Atmospheric Carbon Dioxide*. Retrieved from <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide>

Mail Online. (2019). *Whatever happened to global warming? How freezing temperatures are starting to shatter climate change theory*. Retrieved from

<https://www.dailymail.co.uk/sciencetech/article-1220052/Austria-sees-earliest-snohi-story-America-sees-lowest-temperatures-50-years-So-did-global-warming-go.html>

Meehl, G. A., and Coauthors. (2012): *Climate system response to external forcings and climate change projections in CCSM4*. *J. Climate*, 25, 3661 – 3683. Retrieved from <https://doi-org.proxy.library.nyu.edu/10.1175/JCLI-D-11-00240.1>.

Moore, F. C. 1. fmoore@ucdavis. ed., Obradovich, N., Lehner, F., & Baylis, P. (2019). *Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change*. *Proceedings of the National Academy of Sciences of the United States of America*, 116(11), 4905–4910.

Naomi Oreskes. (2004). *"The Scientific Consensus on Climate Change,"*: Vol. 306 no. 5702 p. 1686 DOI: 10.1126/science.1103618

Nasa Jet Propulsion Laboratory.(2018). *"Ramp-Up in Antarctica Ice Lose Speeds Sea Level Rise"*

Retrieved from <https://www.jpl.nasa.gov/news/news.php?feature=7159>

Pan, Z. (2019). *Climate change and global warming*. Salem Press Encyclopedia of Science. Retrieved from

<http://search.ebscohost.com.proxy.library.nyu.edu/login.aspx?direct=true&db=ers&AN=89475555&site=eds-live>

R. S. Nerem, B. D. Beckley, J. T. Fasullo, B. D. Hamlington, D. Masters and G. T. Mitchum. (2018). *Climate-change–driven accelerated sea-level rise detected in the altimeter era*. DOI: 10.1073/pnas.1717312115

United States Global Change Research Program. (2009). *"Global Climate Change Impacts in the United States,"* Cambridge University Press.