Temperature Trends and Climate Change

Vivian Shu Yi Li, Gianna Saw, Khushi H. Shah, Daniel Veretenov
Northeastern University, Boston, MA, USA

Abstract

The objective of this project was to analyze global warming based on changes in temperature across the United States between the years 1920 and 2010. To this end, a dataset consisting of temperature records from numerous weather stations across the country was utilized. A database was created with multiple tables to store the data in a well-organized manner. Information from the database was extracted and analyzed within a Jupyter Notebook file. Statistical measures such as average temperature, moving average, and regression were used to identify patterns within the data. The project then employed data visualization techniques such as line plots and Folium maps to present and interpret temperature trends across the United States over the span of the 90 years.

The accomplishments of this project demonstrate temperature changes across different regions of the United States. The line plots portray the evolution of temperatures over time. The Folium maps provide a spatial view of temperature variations across different regions of the United States, allowing for analysis of temperature trends shared by different states. The plots created in this project offer a thorough overview of temperature trends over the past century, showing evidence for growing fears of climate change.

Introduction

Climate change and global warming have become highly controversial topics worldwide. Extreme temperatures and climate change have caused lasting damage to the environment and are impacting society. In 2023, scientists reported that approximately 53% of the lakes and around 66% of the reservoirs in the United States have experienced significant water loss, amounting to 22 billion metric tons per year. This loss can be attributed to factors like sedimentation, wildfires, extreme heat, and drought. In contrast, lakes in less populated areas have experienced an increase in storage capacity [1]. In the northern hemisphere, 85% of thousands of glaciers have shrunk, resulting in higher water levels in various places around the world [2]. While scientists have presented compelling evidence of climate change, some individuals remain skeptical. Greta Thunberg, a young Swedish activist, is known for challenging world leaders to take action to mitigate climate change damages [3]. Her campaign “School Strike for Climate” has inspired thousands of young people to organize climate change movements, aiming to improve current environmental policies [4]. Despite her global influence, some people denounced climate activists and denied climate change. In the United States, only 35% of the poll participants claimed to be very concerned about climate change. Meanwhile, 71% of Americans believe climate change is occurring and is related to human activities. The remainder of the population believes that climate change will not affect them or that it is caused by natural causes [5].

Our main research question is “How has climate temperature in the United States changed over the past 90 years?”. We hypothesize that global temperatures are indeed increasing at a concerningly rapid rate based on increasing news of climate catastrophes. The objective of this project is to evaluate the temperature trends over the past decades to gain a better understanding of future trends. By analyzing temperature data from the past decade, we can identify overall patterns in average temperature across the United States and make more accurate predictions about future temperature and climate scenarios in certain regions.

Data Sources and Methods

Data Source

In our study, we obtained our data from data.gov, specifically from the collection of National Oceanic and Atmospheric Administration (NOAA) datasets provided by the Department of Commerce. The dataset we utilized is called the U.S. Daily Climate Normals (1981-2010), which encompasses a 30-year average of weather stations located throughout the United States. The dataset obtained from www.ncei.noaa.gov [6] includes a range of climatic variables, such as temperature, precipitation, wind speed, humidity, and more. While the name of the file states that the data is from 1981 to 2010, the file contains data from as early as 1920 as well.

For our study, we focused specifically on **temperature data.** This dataset contains 90-year (1920-2010) averages of daily temperatures recorded at different weather stations across the United States. By considering such a long time frame, these climate normals provide a reliable baseline for understanding typical climatic conditions in specific regions throughout the year.

Data Cleaning and Integration: Once we sourced the dataset, we proceeded to clean and organize the data to ensure its quality and suitability for analysis. Data cleaning involved several steps, including handling missing values (NaN), removing irrelevant columns, and ensuring data consistency and accuracy. By performing these data cleaning operations, we obtained a refined dataset ready for further analysis.

We then cleaned the data through a series of SQL queries to ultimately produce two tables. The first table contains the station ID and its location as shown below.

	station_id	latitude	longitude	elevation	state	station_name
0	USC00010008	31.5703	-85.2483	139.0	AL	ABBEVILLE
1	USC00010063	34.2553	-87.1814	249.3	AL	ADDISON
2	USC00010140	32.2322	-87.4103	53.3	AL	ALBERTA
3	USC00010160	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
4	USC00010178	33.1272	-88.1550	59.4	AL	ALICEVILLE
...	...	...	...	...	...	...
9749	USW00094992	47.7472	-90.3444	185.9	MN	GRAND MARAIS
9750	USW00094993	45.6689	-96.9914	353.9	SD	SISSETON MUNI AP
9751	USW00094994	43.1561	-90.6775	204.8	WI	BOSCOBEL AP
9752	USW00094995	40.8483	-96.5650	362.4	NE	LINCOLN 8 ENE
9753	USW00094996	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW

The second table, shown below, contains the year and temperature of each month for each station.

	station_id	year	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
0	USC00010148	1894	7.45	6.93	13.26	16.47	20.60	24.62	25.06	25.31	23.10	15.85	8.93	7.72
1	USC00010148	1895	4.05	0.58	10.11	16.48	19.42	24.89	24.61	24.86	23.97	13.57	10.03	5.23
2	USC00010148	1896	4.61	5.82	8.83	18.95	23.39	24.33	26.91	27.49	22.94	15.59	12.19	5.68
3	USC00010148	1897	3.42	8.02	12.05	15.43	19.37	26.45	26.73	25.13	23.34	17.83	9.86	5.96
4	USC00010148	1898	6.44	4.79	12.97	12.74	22.45	26.21	26.21	25.18	23.07	14.78	7.31	4.25
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1134822	USW00094996	2006	3.31	-0.89	4.08	13.41	17.62	22.99	26.20	23.72	16.77	10.26	5.06	1.46
1134823	USW00094996	2007	-4.16	-4.41	8.73	10.02	18.10	21.31	25.11	25.66	18.95	13.85	4.59	-4.18
1134824	USW00094996	2008	-5.21	-3.60	3.56	8.05	15.06	21.03	24.18	22.45	18.12	12.13	4.50	-4.87
1134825	USW00094996	2009	-4.01	0.09	4.52	9.87	16.92	21.46	21.95	21.59	17.61	7.82	7.60	-6.96
1134826	USW00094996	2010	-7.58	-5.42	5.21	13.24	15.39	22.47	24.65	24.82	18.68	13.76	4.19	-3.45

By merging these two tables based on the station ID, we created a comprehensive database that formed the basis for our analysis.

	station_id	year	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	latitude	longitude	elevation	state	station_name
0	USC00010160	1893	3.56	11.46	11.83	19.24	20.06	23.96	27.19	25.95	22.96	16.77	11.43	9.05	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
1	USC00010160	1894	8.63	8.26	14.63	18.10	21.71	24.46	25.42	25.51	23.29	17.02	10.33	9.16	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
2	USC00010160	1896	6.04	8.55	11.24	20.17	23.77	24.83	26.64	27.45	24.17	17.11	14.30	7.72	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
3	USC00010160	1897	5.55	10.53	15.45	16.78	20.45	27.39	27.15	25.54	23.23	18.96	12.23	8.62	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
4	USC00010160	1898	9.28	7.46	15.18	14.70	22.64	26.63	26.41	25.25	23.32	16.15	9.77	6.43	32.9453	-85.9481	195.1	AL	ALEXANDER CITY
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
756557	USW00094996	2006	3.31	-0.89	4.08	13.41	17.62	22.99	26.20	23.72	16.77	10.26	5.06	1.46	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW
756558	USW00094996	2007	-4.16	-4.41	8.73	10.02	18.10	21.31	25.11	25.66	18.95	13.85	4.59	-4.18	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW
756559	USW00094996	2008	-5.21	-3.60	3.56	8.05	15.06	21.03	24.18	22.45	18.12	12.13	4.50	-4.87	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW
756560	USW00094996	2009	-4.01	0.09	4.52	9.87	16.92	21.46	21.95	21.59	17.61	7.82	7.60	-6.96	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW
756561	USW00094996	2010	-7.58	-5.42	5.21	13.24	15.39	22.47	24.65	24.82	18.68	13.76	4.19	-3.45	40.6953	-96.8542	418.2	NE	LINCOLN 11 SW

756562 rows × 19 columns

Methods

Our report uses a combination of quantitative and qualitative analysis. Our analysis primarily focuses on time series analysis, regression modeling, and data visualization using Matplotlib and Folium. By combining qualitative and quantitative analysis, we gained a comprehensive understanding of the temperature trends in the US.

First, **time series analysis** was used to explore temperature trends over time. This method allows us to uncover patterns, trends, and seasonality within the temperature data. By examining the comprehensive database, we identified long-term temperature patterns for each station. Initially, line plots were used to visualize temperature variations over the years, but due to cluttered plots, we turned to regression analysis.

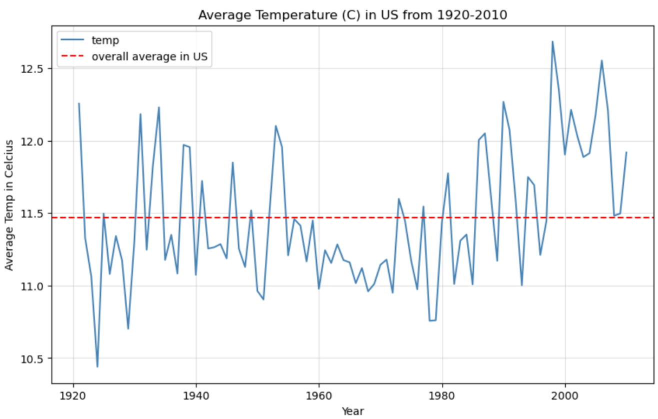
To increase the legibility of the plots and to better understand the temperature trends, we performed **regression analysis** for each state. Regression models enable us to estimate the average temperature change per year and identify the overall trend. By fitting regression lines to the temperature data, we gained insights into the direction and magnitude of temperature changes over time in different states. This quantitative analysis provided a clear depiction of the long-term temperature trends and allowed for a more concise representation of the data.

In order to present our findings effectively, we employed **qualitative data visualization** techniques using **Matplotlib and Folium**. Matplotlib, a versatile Python library, allowed us to create plots such as line graphs and regression plots. These visualizations allowed us to interpret temperature trends and helped to identify any significant deviations or patterns.

Furthermore, we utilized **Folium**, a mapping library, to create interactive maps that displayed the spatial distribution of temperature variations across different locations or stations. By overlaying the temperature data on maps, we were able to visualize the geographical patterns of temperature changes, thereby allowing us to portray temperature changes across multiple states and regions.

Analysis

Graph 1: Line Plot of the Average Temperatures in the US from all the Station IDs (1920-2010)



The graph displays a significant increase in temperatures between the 1940s and 1950s, likely due to the post-World War II industrial boom. This period experienced rapid economic growth and industrial development, leading to skyrocketing emissions of greenhouse gases and aerosols from factories, power plants, and more. These emissions contributed to rising temperatures across the Earth's surface. [7] Furthermore, a significant temperature drop below the average (shown in red) was observed across all the station IDs before the early 1980s. This decrease aligns with the implementation of environmental regulations in the 1970s and 1980s targeting aerosol pollution reduction. Aerosols reflect the sun's rays away from the Earth's surface, thus contributing to cooling temperatures [8].

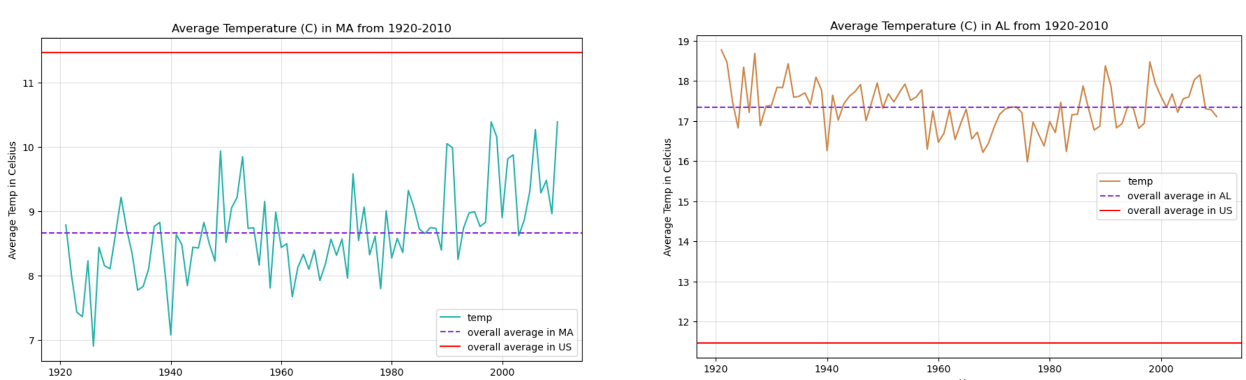
After the 1980s, increasing fluctuations in temperature were observed. The overall temperature in the United States trended upwards above the average temperature line (shown in red on the graph).

	state	average_slope
16	MA	0.015129
36	RI	0.014560
25	ND	0.013833
5	CT	0.013823
28	NJ	0.013399
18	ME	0.013182
3	CA	0.012656
27	NH	0.012159
20	MN	0.012062
19	MI	0.011729

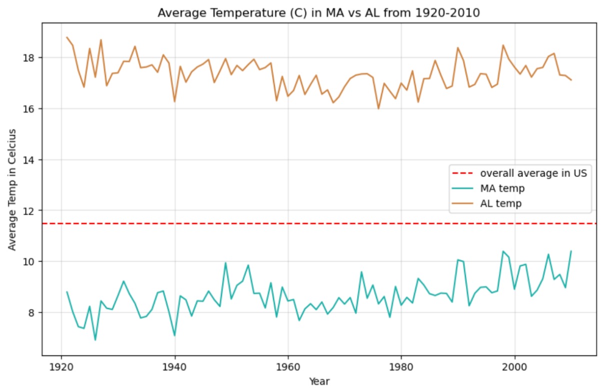
Regression Analysis - Regression analysis was performed to better understand average temperature change per year and to identify the overall trend. The top 10 states that recorded the highest number of changes in temperature are shown on the left.

Massachusetts was the state that was most affected with a moving average slope of 0.015129. However, Alabama ranked last, with a slope of -0.002, meaning that it ranked consistently in terms of changing temperature average values. To better understand the difference, we plotted the average values of Massachusetts and Alabama and their overall average temperature from 1920-2010.

Graphs 2 & 3: Average Temperatures in MA and AL from 1920-2010



Graph 4: Combining the Average Temperature in Massachusetts and Alabama from 1920-2010



Typically, Alabama experiences higher temperatures than Massachusetts due to its geographical location. However, specific years like the 1950s saw an increase in temperatures. The temperature graph for Alabama shows a consistent decrease in temperature. In contrast, Massachusetts demonstrates a general upward trend, indicating increasing temperatures. This pattern suggests that hotter regions like Alabama are gradually becoming cooler, while colder regions like Massachusetts are gradually becoming warmer.

Considering the post-2010 temperature trends, we can consider that the temperatures in Alabama and Massachusetts might eventually converge and become comparable. To better visualize the results, we plotted our regression results through Folium maps (Graph 5).

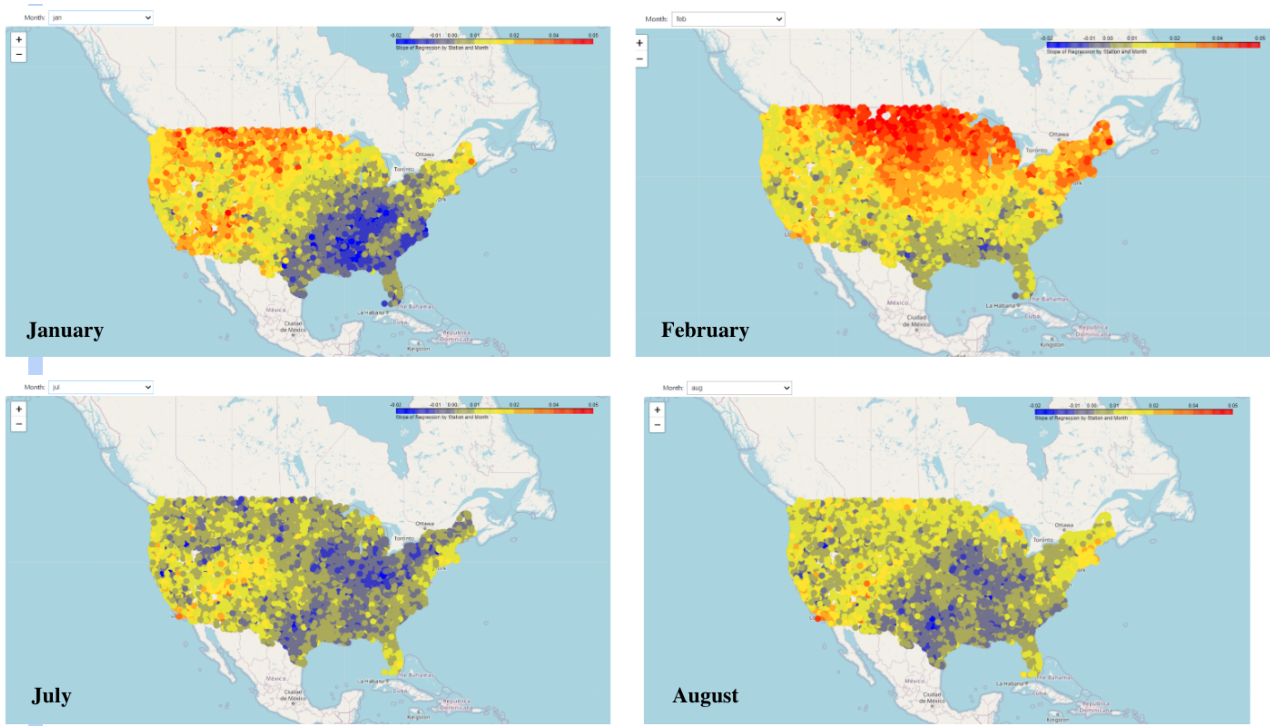
In July and August, the map displays predominantly cooler colors such as blue and green, indicating a negative regression. This suggests a decreasing trend in temperatures during these months. When examining the visual differences between the two months, particularly in the Southern region, we notice a shift from blue dots to yellow dots. This change in color reflects a slight warming trend and denotes a temperature variation.

This visual contrast between blue and yellow dots suggests that there is some level of temperature change occurring within a month. While the colors remain on the cooler end of the spectrum, the presence of yellow dots indicates a warming trend from the previous cooler temperatures.

The maps of January and February exhibit a noticeable and rapid change in average temperature, particularly in the Midwest and South regions. In January, the Midwest region is predominantly represented by yellow dots, with some signs of red dots. However, within the span of only one month, the region experiences a significant shift, becoming dominated by red dots in February. This shift indicates a stronger regression or temperature change during this period.

A similar trend can be observed in the South region of the map. In January, this area is mainly characterized by blue dots, indicating cooler temperatures. However, as February arrives, there is an increase in the number of yellow dots, portraying a slight increase in temperature. The visual contrast observed between the maps of January and February suggests that these months experience dynamic temperature fluctuations, which may be influenced by various factors such as weather systems or regional climate patterns. Below are side-by-side visual comparisons of temperatures changes (in Celsius) across the United States for four different months for years 1920-2010:

Graph 5: Regression Folium Maps Over 90 Years (*Left to right: Jan, Feb, July, August)*



Conclusion:

Overall patterns in average temperature across the United States appear to vary by region. For example, states in the Northeast appear to have increasing average temperatures, while states in the Southeast appear to have decreasing average temperatures. However, despite this apparent disparity between regions, nearly all regions of the country are seeing increasing temperatures, with only some states seeing slightly cooler temperatures. Interestingly, Massachusetts showed the highest change in temperature (demonstrated by having the highest slope of regression), while Alabama showed the lowest.

Based on the results of this project, we predict that temperature trends will largely continue as they are. With the trend of typically-warmer states (such as Alabama) becoming cooler and typically-cooler states (such as California) becoming warmer, we can cautiously entertain the possibility that these two states' temperatures may converge one day. However, given the overall trend in temperatures since 2010, coupled with alarming research from scientists who study climate change, it can be reasonably predicted that temperatures overall will show an increasing trend. These trends in temperature can contribute to climate change through avenues such as rising sea levels, biodiversity loss of species whose habitats require a certain temperature, and more.

This project's accomplishments include the creation of plots and maps that effectively portray changes in temperature across time and regions of the United States. By utilizing a dataset collected from numerous weather stations across the country, the project was able to provide insights into temperature patterns and variations over the span of the 90 years. Through the creation of line plots, we were able to reveal temperature trends and fluctuations, which allowed us to identify significant changes and patterns. The line plots also allowed us to conduct an investigation of historical events that may have contributed to temperature fluctuations. Through the creation of Folium maps, we were able to understand the geographical distribution of temperature changes, allowing us to identify regions with unique or shared temperature patterns. These maps offered a visual representation of temperature variations across different regions and states in the United States.

The results of this project can provide valuable insights for researchers, lawmakers, and scientists who are looking to combat the effects of climate change. However, it is also important to note the limitations of this work. Primarily, climate change is caused by a plethora of factors that interact with each other in numerous ways. The dataset that was used for this project only contained data for one factor: temperature. While the findings of this project are significant, further research and analysis should focus on other important factors such as human activities and deforestation to analyze trends in climate change. Furthermore, since this project focused on the years 1920 to 2010, future analysis for years beyond 2010 could be invaluable to understanding recent temperature trends and to determining if climate change is accelerating.

Project Contributions

The project team consisted of Daniel Veretenov, Vivian Li, Khushi Shah, and Gianna Saw. Daniel Veretenov played a crucial role in the coding aspect of the project. His proficiency in programming languages, including Python and SQL, allowed for the complex analysis of the dataset. Vivian Li contributed primarily to the report writing and the presentation slide. She also assisted in the coding aspect of the project. She worked on summarizing and implementing research methodology and background. Khushi Shah's primary focus was on the report writing and editing phase of the project. She worked on analyzing the plots that were created and effectively summarizing the research methodology, results, and implications. Gianna Saw also played a significant role in the report writing process. She worked on analyzing the methodology and results of the project, while also providing insights into the analysis conducted.

References

[1] Paddison, L. (2023, May 18). *The world's largest lakes are shrinking dramatically, and scientists say they have figured out why*. CNN. Retrieved June 21, 2023, from https://www.cnn.com/2023/05/18/world/disappearing-lakes-reservoirs-water-climate-intl/index.html

[2] Hersher, R. (2022, February 8). *Climate change spells disaster for the world's glaciers*. NPR. Retrieved June 21, 2023, from https://www.npr.org/2022/02/08/1079227805/glaciers-are-shrinking-fast-scientists-are-rushing-to-figure-out-how-fast

[3] Thunberg, G. (2022, October 8). *Greta Thunberg on the climate delusion: 'We've been greenwashed out of our senses. It's time to stand our ground'*. The Guardian. Retrieved June 21, 2023, from https://www.theguardian.com/environment/2022/oct/08/greta-thunberg-climate-delusion-greenwashed-out-of-our-senses

[4] Kraemer, D. (2021, November 5). *Greta Thunberg: Who is the climate campaigner and what are her aims?* BBC. Retrieved June 21, 2023, from https://www.bbc.com/news/world-europe-49918719

[5] Doherty, E. (2022, August 15). *New poll shows American doubts on personally affecting climate change*. Axios. Retrieved June 21, 2023, from https://www.axios.com/2022/08/15/poll-climate-change-americans-congress

[6] *U.S. Daily Climate Normals (1981-2010) - Catalog*. (n.d.). Data.gov. Retrieved June 22, 2023, from

https://www.ncei.noaa.gov/pub/data/normals/1981-2010/source-datasets/coop.52g.20110530.FLs.52g.tar.gz

[7] Pruitt, S. (2020, May 14). *The post World War II Boom: How America Got Into Gear*. History.com. https://www.history.com/news/post-world-war-ii-boom-economy

[8] *The discovery of global warming*. Niels Bohr Library & Archives. (2023, May). https://history.aip.org/climate/20ctrend.htm