Climate Data Analysis in America

(COMP3125 Individual Project)

William Weinberg
Wentworth Institute of Technology,
Data Science

***Abstract*— This project analyzes data of three important aspects of weather inside the United States: temperature change, precipitation patterns, and tornado magnitude over time. Areas from the west coast, midwest, and east coast were selected in order to have a fair sample size. By combining data analysis and visualizations it is possible to provide insights into the evolving climate patterns in the United States, and it can be found that temperature is increasing, rain is increasing, and magnitude of tornadoes are decreasing.**

Keywords—climate, United States, temperature, precipitation, tornados

I. Introduction (Climate Change and Its Impact)

Climate change is a pressing global issue with complex consequences that affect humans, animals, and the environment itself. Understanding climate patterns is integral in order to see what the future holds and minimize any possible risks. This project focuses on three aspects of climate change in the United States: temperature, precipitation, and tornado activity. Temperature is a fundamental variable when it comes to discussing climate and its impact on human life, ecosystems, and economies. By analyzing temperature trends in different regions of the US, we aim to identify any patterns or potential causes of these changes. Precipitation is important as it is critical to agriculture and water resources. Studying rates of precipitation can allow us to predict future flood risks. Tornados are severe weather events that can cause significant harm and lead to injury or loss of life. By reviewing historical tornado data and exploring tornado intensity we can begin to understand what the future holds and tailor mitigation strategies or increase disaster preparedness. By investigating these three variables of climate change we can gain valuable insights to the evolving climate of the United States and better understand what lies ahead.

II. Datasets

A. *Source of dataset (Research)*

This project utilized three publicly available datasets from reputable sources to analyze climate patterns within the US.

Dataset 1: United States Rainfall

Source: Kaggle (https://www.kaggle.com/datasets/iamkevin/united-states-rainfall)

Originator: National Climate Data Center (NCDC) (https://www.ncdc.noaa.gov/cdo-web/)

Credibility: The NCDC is a leading government agency within the National Oceanic and Atmospher Administration (NOAA), known for its reliable and long-term climate data collection. Using data curated by the NCDC ensures accuracy of the rainfall information.

Dataset 2: Monthly Mean Temperature Data for Major US Cities

Source: Kaggle (https://www.kaggle.com/datasets/garrickhague/temp-data-of-prominent-us-cities-from-1948-to-2022

Originator: Climate Prediction Center (CPC) (https://www.cpc.ncep.noaa.gov/

Credibility: The CPC is another branch of the NOAA dedicated to weather and climate prediction. Their data is widely used by scientists due to its high quality and established methods for collecting data.

Dataset 3: Tornados in North America

Source: Kaggle (https://www.cpc.ncep.noaa.gov/

Originator: National Weather Service's Storm Prediction Center (SPC) (https://www.spc.noaa.gov/)

Credibility: The SPC, a division of the National Weather Service, is a trusted source for severe weather data such as tornados. Their data collection procedures ensure accuracy for analysis.

B. *Character of the datasets*

Dataset 1: The dataset on US Rainfall has a total of 899 entries. It consists of columns which include station identification number, state code, station list number, network division number, element units, the date, hour, amount of rainfall, and state. The main columns used were date, amount of rainfall, and state.

station_id	state_code	station_list_no	network_div_no	element_units	date	hour	rainfall	state
22079200	22	792	0	HI	1955-12-03	2500	00297	Mississippi
22079200	22	792	0	HI	1955-12-04	2500	00022	Mississippi
22079200	22	792	0	HI	1955-12-05	2500	00009	Mississippi
22079200	22	792	0	HI	1955-12-06	2500	00026	Mississippi
22079200	22	792	0	HI	1955-12-18	2500	00083	Mississippi
22079200	22	792	0	HI	1956-01-01	2500	00000	Mississippi

Dataset 2: The dataset on US City Monthly Temps has a total of 269657 entries. It consists of a column on time and each state has its own column detailing the monthly temp. The main columns used were time, Boston, Boise, and Dallas.

time	albuquerqu	anchorage	atlanta	boise	boston	buffalo	charlotte	chi
1948-01-01	-0.4400024	-5.97998053	4.200134	-0.6000061	-5.98999	-7.79000852	2.3900146	-7.3
1948-02-01	1.6499939	-9.0	9.110016	0.13000488	-4.350006	-4.549988	6.6700134	-3.1
1948-03-01	3.4700012	-5.850006	13.640015	2.3099976	2.4100037	1.1600037	12.52002	2.9
1948-04-01	13.240021	-0.269989	18.660004	7.25	8.130005	8.690002	17.279999	12.
1948-05-01	16.910004	6.730011	21.540009	11.860016	12.77002	11.320007	20.619995	13.
1948-06-01	21.040009	11.040009	25.640015	17.850006	17.360016	17.580017	24.830017	20.
1948-07-01	24.230011	12.080017	26.410004	19.709991	22.350006	21.190002	26.820007	23.
1948-08-01	23.330017	10.730011	25.140015	19.140015	21.929993	20.119995	24.890015	23.
1948-09-01	19.709991	6.9200134	22.080017	14.410004	17.51001	17.230011	21.190002	19.
1948-10-01	12.130005	2.1900024	15.679993	8.830017	10.880005	9.0	14.429993	10.
1948-11-01	2.7700195	-8.290009	13.110016	1.5700073	8.459991	7.2700195	12.869995	6.3
1948-12-01	1.8600159	-12.4499829	9.149994	-5.0899963	0.8200073	-0.269989	8.050018	-0.

Dataset 3: The dataset on Tornados in North America has a total of 57987 entries. It consists of columns which include origin, year and date, time zone, state, state code, state name, magnitude, injuries and fatality, loss and crop loss, and longitude/latitude. The main columns used were magnitude and date.

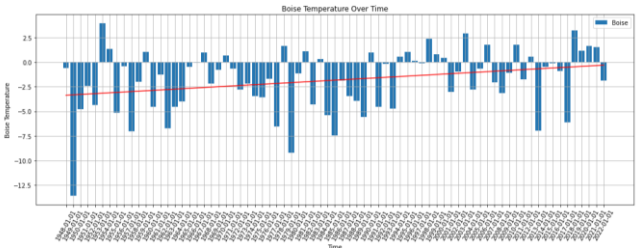
979-8-3315-0993-4/25/$31.00 ©2025 IEEE

FID	OM	YR	MO	DY	DATE	TIME	TZ	ST	STF	STN	MAG	INJ	FAT	LOS
49942	365283	2012	4	14	2012-04-14	20:22:00	3	IA	19	0	1	0	0	0.07
49943	411	1982	5	17	1982-05-17	16:32:00	3	MN	27	6	2	3	0	6.0
49944	372581	2012	4	14	2012-04-14	20:35:00	3	IA	19	0	1	0	0	0.0
49945	410	1982	5	17	1982-05-17	16:45:00	3	MN	27	5	2	0	0	4.0
49946	383794	2012	5	3	2012-05-01	19:40:00	3	IA	19	0	2	0	0	1.0
49947	412	1982	5	17	1982-05-17	17:00:00	3	MN	27	7	1	0	0	4.0

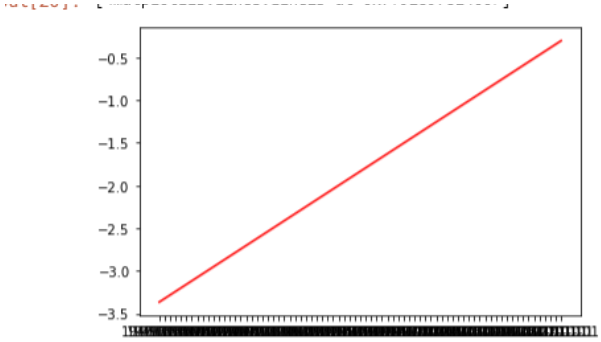
III. Methodology

To ensure accurate analysis, the raw datasets were filtered to include only relevant datapoints. For instance, temperature data was filtered for Boston, Boise, and Dallas as this sample size represents the three major regions of the USA. Similarly, rainfall data was filtered for Washington, Oklahoma, and Massachusetts. It was also sorted chronologically. To analyze long term trends, linear trendlines were fitted to the data. This was used to identify upwards or downwards trends in the long term. Python libraries that were used to conduct analysis and visualizations were Pandas, NumPy, and Matplotlib. By combining these tools, we can draw reasonable conclusions about the data.

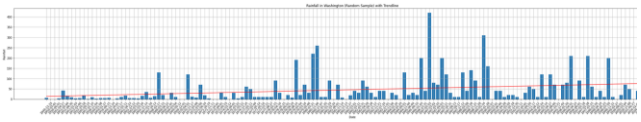
IV. Results



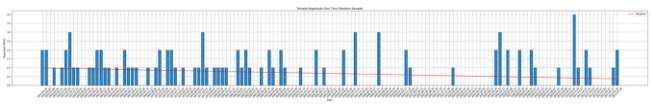
Looking at the temperature analysis for the west coast, midwest, and east coast there is a clear upwards trend in average temperatures This trend is consistent with global climate change patterns and suggests a warming climate in the United States.



Between the three regions, temperatures have gone up between 1.2 to 3 degrees warmer.



The precipitation analysis shows that there is also an upwards trend between the three regions, albeit not as strong. Washington shows this trend quite well as it goes from on average 25 inches of rainfall to up to 60. This is also consistent with global climate change as rainfall is expected to increase by 0.03 inches per year.



Tornado magnitude in America seems to trend down as time goes on. It begins with an average of 1.0 magnitude and lowers to approximately 0.4. There are multiple reasons behind this such as better construction and disaster preparedness. Changes in atmospheric circulation patterns in different areas could also explain this.

V. Discussion

There are limitations such as the data. The climate data that dates to the 1950s could be less accurate than modern data due to technological advances and gaps in data. The use of trend analysis is very useful in displaying long term trends but might not capture nonlinear relationships. The sample sizes used in the analysis may impact the statistical signifigance of the findings. Some of the data sets have hundreds of thousands of entries and only so much can be used to create a visualization that is still legible. Larger sample sizes and more robust stat tests could strengthen the conclusions. Future research could include deeper regional analysis than just midwest, west coast, and east coast. Data from other fields such as agriculture and ecology could also be cross referenced to investigate the impact of these climate factors.

VI. Conclusion

This project aimed to analyze key aspects of climate change in the US, focusing on temperature, precipitation, and tornado activity. By analyzing historical data from reliable sources, we were able to identify significant trends and patterns. These include: a general upward trend in temperatures which was observed across various regions of the US. This trend is consistent with global climate change patterns The analysis of precipitation showed increased rainfall consistent with scientific research showing that rainfall increases every year on average. This could impact water resources and agriculture. Tornado data suggested a potential downward trend in magnitude, though it is important to note that tornado activity can be influenced by various other factors. These implications highlight the urgent need for climate change mitigation strategies to reduce greenhouse gas emissions and limit global warming. Understanding these patterns could help develop policies to minimize extreme weather events or damage to ecosystems.

References

[1] National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCDC). (n.d.). Climate Data Online https://www.ncdc.noaa.gov/cdo-web/

[2] National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC). (n.d.). Climate Prediction Center. https://www.cpc.ncep.noaa.gov/

[3] National Oceanic and Atmospheric Administration (NOAA) Storm Prediction Center (SPC). (n.d.). Storm Prediction Center. https://www.spc.noaa.gov/