A Comprehensive Analysis of Temperature Trends in China: Examining the Evidence for Temperature Increase

December 14, 2023

JC3007: Scientific Research Methods (2023-2024)
Assessment: A2. Project Report

Student Name: Hou Yifei
ID: 50079704

University of Aberdeen

1 Introduction

In daily life, you can feel that this winter is warmer than last year or that summer is hotter. Climate change is increasingly affecting our lives. Research shows that in 2021, the average temperature of the global ocean surface has increased by 1.21 °C compared to the average temperature in the pre-industrial era (1850-1900)[1].

Climate change constitutes a vast field of research. My goal is to verify the phenomenon of global climate warming through detailed research work and experience a real scientific research process first-hand. This way, I can have a more solid and in-depth argument when advocating for planetary protection in the future.

In this study, I analyzed temperature data from Chinese meteorological stations from January 1, 1980 to December 31, 2022. The purpose is to study temperature trends at various geographical levels.

The research aims to achieve two core objectives:

1. Demonstrate the trend of temperature changes in various geographical levels in China through clear statistical analysis, as a side evidence to support the phenomenon of global climate warming.

2. Reveal temperature change patterns across China's provinces and administrative regions, providing important insights and valuable data to policymakers and other scholars.

2 Related Work and Background

The first step in my research work is to understand the current scientific research on climate change, to understand the challenges faced by mankind and the achievements it has made, as well as the direction of current scholars' efforts. The research I have to do is small, but knowing the big picture helps me know where to go.

After understanding the big picture of climate change research, I need to understand the methods required for my small research and develop an appropriate study plan to master the research methods.

2.1 Key Research Areas of Climate Change

The Intergovernmental Panel on Climate Change (IPCC) is an international organization established in 1988 by the United Nations Environment Program (UNEP) and the World Meteorological Organization (WMO). It evaluates scientific research on climate change, its impacts and potential future risks, and proposes adaptation and mitigation strategies. The IPCC is best known for its comprehensive assessment reports, which are the leading international scientific consensus on climate change.

In the IPCC's latest research report, The IPCC's Sixth Assessment Report, the main areas of current climate change research are summarized.[2] These areas are:

1. Attribution of Climate Change: Understanding the role of human influence on climate system changes and how it intensifies specific weather events like heatwaves and heavy rainfall.

2. Regional Climate Change Impacts: Assessing how climate change affects different regions, resulting in varied changes such as increased precipitation in high latitudes and intensified water cycles leading to floods and droughts.

3. Mitigation Strategies: Research ways to reduce greenhouse gas emissions to stabilize climate, with a clear understanding that CO_2 is the main driver of climate change, and that efforts to limit other pollutants, such as methane, could also benefit health and climate.

4. Long-term Projections: Providing detailed assessments of potential future changes, including sea level rise, changes in ocean systems, and loss of Arctic sea ice, all of which will continue to advance with further warming.

2.2 Analyzing Temperature Trends: Methods and Practices in Climate Research

The National Center for Atmospheric Research (NCAR), located in the United States, specializes in atmospheric and Earth system sciences. Engaged in diverse research fields including climate change, meteorological forecasting, air quality, and space weather, NCAR is dedicated to understanding atmospheric behaviors and interconnected environmental systems. Its website serves as a rich repository, offering a plethora of research findings, datasets, and educational materials, thus becoming an essential resource for scientists, educators, and anyone interested in the field of atmospheric sciences.

On the NCAR website, I found a method to analyze temperature trends, crucial in climate research for identifying, estimating, and forecasting trends. A common method used is simple linear regression, a statistical approach to analyze the linear relationship between two variables, like temperature over time. This involves finding a regression line (‘y = ax + b‘) that minimizes the sum of vertical distances from all data points to this line, where ‘y‘ is the dependent variable (temperature), ‘x‘ is the independent variable (time), ‘a‘ is the slope, and ‘b‘ is the intercept. Data collection, such as gathering average temperatures for several years, is essential for practical implementation, often using statistical software like R or Python. To assess the statistical significance of these trends, methods like the Student-t test and the Mann-Kendall test, which is less sensitive to outliers and skewed distributions, are used. This comprehensive approach aids in understanding and validating temperature trends over time.

2.3 Selecting Statsmodels for In-Depth Data Analysis

Python has many extensive mathematical analysis libraries, of which Scikit-learn, statsmodels, and SciPy are frequently used in the industry. All three libraries provide powerful linear regression analysis capabilities. Scikit-learn is the first choice for machine learning, but it does not provide P-Value calculation for hypothesis testing. Both statsmodels and SciPy can provide statistical analysis, but statsmodels provides a more comprehensive set of statistical model analysis tools, including detailed output and advanced statistical tests, making it ideal for in-depth statistical inference and reporting. After careful consideration, statsmodels was selected as the data analysis library for this study.

3 Research Question

The research encompassed a comprehensive process: data collection from GHCN[3], rigorous validation using Python, and analysis via statistical methods. Statsmodels[4] will be used for linear regression and hypothesis testing to confirm trends. The findings are presented through insightful charts and interpretive data analysis.

3.1 Obtaining Historical Weather Data

We will use the Global Historical Climatology Network (GHCN) dataset[3], a comprehensive climate data collection including temperature and precipitation, maintained by the National Climatic Data Center (NCDC). This dataset is critical for analyzing long-term climate trends and is applied in various research and policy-making fields. For this study, data for each provincial capital in China from January 1, 1980, to December 31, 2022, will be downloaded from GHCN.

3.2 Perform data verification

After obtaining the historical weather data, Python will be used to verify its accuracy and completeness. The process includes reading the downloaded data for each meteorological station and using a graphical representation, with the X-axis being the year and the Y-axis being the temperature. An initial inspection is performed by the human eye. Visualizing the data helps identify any anomalies or outliers, which is critical in determining the need for data cleaning. This step ensures the integrity of our analysis, as clean and accurate data is the required basis for drawing reliable conclusions about climate trends.

3.3 Hypothesis testing for temperature trends

To scientifically validate the hypothesis of temperature increases across various geographic levels, from cities to larger regions, we define a null hypothesis (H0) and an alternative hypothesis (H1). To scientifically articulate the hypothesis regarding a city's temperature increase, we establish a null hypothesis (H0) and an alternative hypothesis (H1):

1. H0 (Null Hypothesis): There is no significant change in the annual average temperature across specific geographic regions, suggesting that observed temperature changes are random and any identified trends occur by chance.

2. H1 (Alternative Hypothesis): There is a significant trend, whether upward or downward, in the annual average temperature across specific geographic regions, suggesting that the observed changes in temperature are not mere coincidences but indicative of real climatic shifts.

3.4 Statistical analysis of temperature trends

In the analysis using the statsmodels library[4], linear regression was applied to temperature data. We calculated the slope of the regression line, representing the rate of temperature change. For statistical rigor, the P-Value associated with null hypothesis (H0: no significant temperature change) was computed. This P-Value measures the probability of observing the data if H0 were true. A P-Value below a specific threshold (commonly 0.05) indicates statistically significant evidence

against H0, suggesting a real temperature change. If the regression slope is positive and the P-Value is sufficiently low, it denotes a significant upward or downward temperature trend.

3.5 Statistical analysis of temperature trends

The final analysis results are presented across four levels: meteorological stations, provinces, administrative regions, and the national level. Each level has a corresponding results table, detailing the statistical outcomes for each geographic area. These tables allow us to discern the trends and magnitudes of temperature changes in each geographic area. The form of the table is as Table1

Provinces	Regression Slope	P-value	Temp. Rise
Beijing			TRUE
Guangdong			

Table 1: Temperature Trend Analysis of Provinces

The most effective way to display these variations is by visually representing the temperature changes on maps, highlighting the spatial distribution and intensity of climate shifts across different regions. This approach not only makes the data more accessible but also provides a clearer understanding of the climatic changes occurring within China.

4 Experimental Design

The experimental design of this study focuses on systematically quantifying and visualizing long-term temperature trends in China, centered around testing the null hypothesis (H0) that there is no significant long-term temperature trend in China. By using a complete and reliable data set, the study aims to accept or reject H0 through stages of data cleaning, statistical analysis and clear visual presentation. These processes are driven by a series of Python scripts, each tailored to a specific research goal. The analysis effectively illustrates temperature trends across China through detailed statistical data and comprehensive visual presentations.

All code and replication methods for this project are available on GitHub. Access the full project details and resources at JC3007 Project GitHub repository

4.1 Data Collection and Preparation

The National Oceanic and Atmospheric Administration (NOAA) is the premier provider of a wide range of climate data, including reliable and comprehensive temperature records. Downloading weather data from NOAA is a smart choice. The date range of the data is January 1, 1980, to December 31, 2022. During this period, China has established a complete meteorological observation system. The data during this period is complete and has analytical value. Download data for all meteorological stations across China from the NOAA database. These data can fully reflect the temperature changes in China's vast territory.

Select CSV format when downloading to facilitate processing by python scripts.

4.2 Data Processing

The data downloaded from NOAA must be processed to ensure the accuracy and reliability of the analysis by correcting, standardizing and optimizing the original data sets.

The data processing procedure is as follows:

1. **Dataframe Creation:** Reading all CSV files into a pandas DataFrame.
2. **Data Cleaning:** Removing incomplete rows, such as those with incorrect dates or missing temperature values.
3. **Station Data Validation:** Data from a meteorological station is considered invalid and removed from the DataFrame if it does not meet the following criteria:
    - The station must have data for every month from January 1, 1980, to December 31, 2022.
    - Each month must have more than 25 days of data.
4. **Temperature Conversion:** Converting all temperature readings to Celsius.
5. **Data Visualization:** Visualizations display year (X-axis) against temperature (Y-axis), including average, maximum, and minimum temperatures, to assess data completeness.

4.3 Temperature Trend Analysis

The analysis focuses on temperature trends across China, from local meteorological stations to national trends. It involves:

- Adding geographic information for province, region, and administrative code.
- Applying OLS Regression to annual average temperatures at each geographic level.
- Storing results in CSV files for further analysis, including slope and P-Value from the OLS model.

4.4 Visualization Design

To represent China's temperature trends, two key visualizations are created:

1. **Meteorological station-level map:** Showcasing temperature changes at each station to elucidate local climate variations.
2. **Provincial-Level Map:** Summarizing station data within provinces to depict broader regional trends.

Both maps utilize a color gradient from dark blue (significant temperature drops) to dark red (major temperature increases), providing an intuitive understanding of annual temperature changes across different scales.This design ensures a thorough and accurate representation of temperature trends, enabling acomprehensive understanding of climatic changes in China.

4.5 Expected Outcomes

This study aims to ultimately verify or refute the null hypothesis and elucidate temperature trends in China. Expected results include:

- **Statistical Validation:** Determine the statistical significance of long-term temperature trends in China based on OLS regression P-Values.
- **Quantitative Insights:** Revealing the magnitude and direction of temperature changes from local meteorological stations to national scales.
- **Visual Interpretation:** Maps clearly illustrate temperature trends and highlight key areas of climate change.

5 Results

After rigorous data cleaning, the study performed OLS analysis and the analysis results were saved in three CSV files. The files can be viewed on JC3007 GitHub repository, and this section outlines the key analysis results.

5.1 Data Validity

Of the 226 meteorological stations across China for which data were downloaded, 139 stations met the criteria for data validity. This criterion necessitated that each station have data for every month from January 1, 1980, to December 31, 2022, with over 25 days of data per month. Figure 1 illustrates the proportion of stations meeting the data validity criteria.

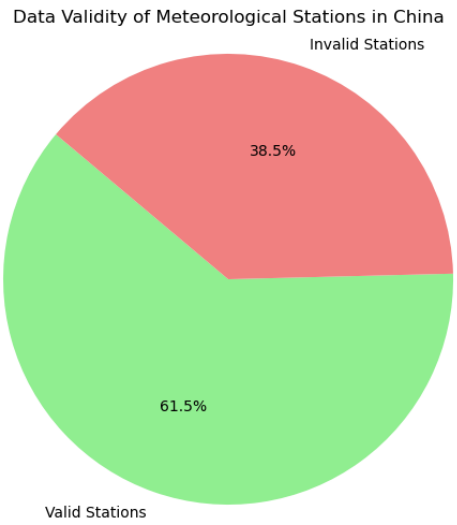

Figure 1: Proportion of meteorological stations meeting the data validity criteria.

5.2 Visual Analysis

The results of the study are made readily apparent through intuitive visualizations of temperature changes on maps of China, segmented by meteorological stations and provinces.

5.2.1 Meteorological Station-Level Map Analysis

The meteorological station-level map, as shown in Figure 2, reveals the following insights:

- The meteorological stations for analysis are distributed throughout China, and the data from the meteorological stations can reflect the meteorological conditions across China.
- The majority of meteorological stations indicate an increasing trend in temperature, reflecting a widespread pattern of warming.

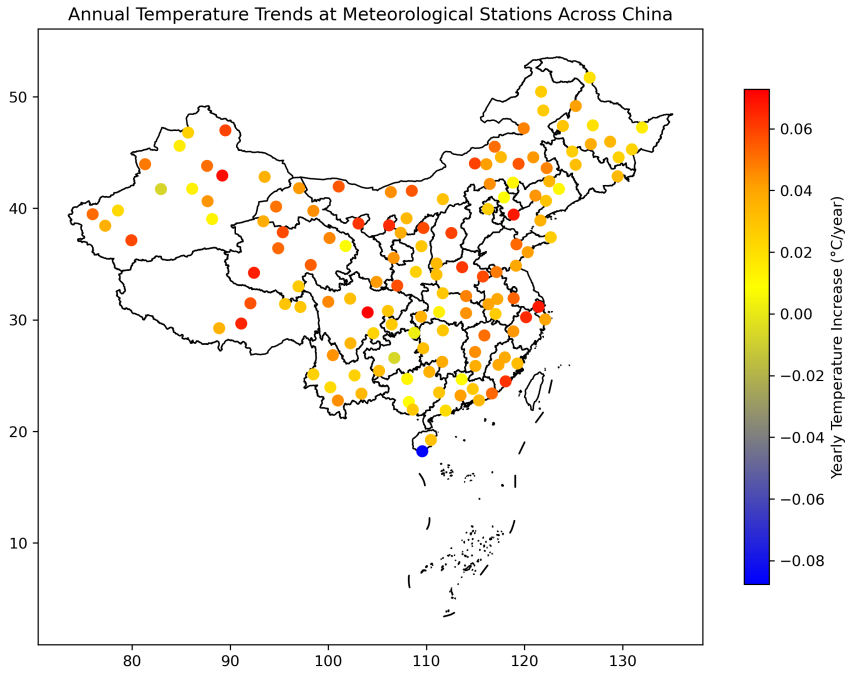

Figure 2: Annual Temperature Trends at Meteorological Stations Across China

5.2.2 Provincial-Level Map Analysis

Analysis of the provincial-level map, as depicted in Figure 3, yields these conclusions:

- Most provinces are experiencing a rise in temperature.
- The extent of temperature increase in the southern provinces is less pronounced compared to other regions.
- When correlating with topographical maps of China, it is observed that mountainous areas in the south show a smaller increase in temperature, while the central and eastern plains exhibit a larger increase.

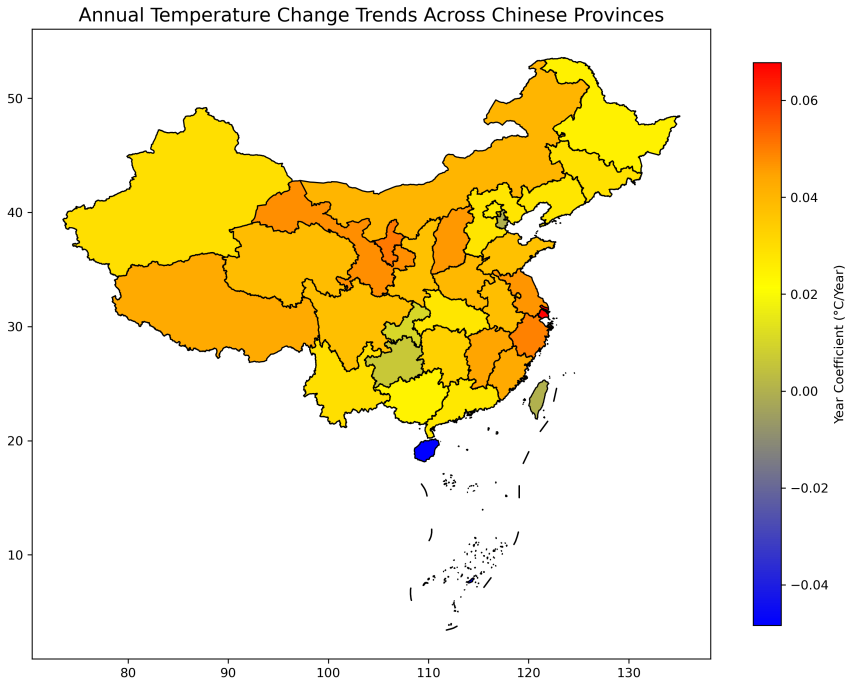

Figure 3: Annual Temperature Change Trends Across Chinese Provinces

5.3 Quantitative Analysis

The data reveal significant insights into the climatic changes occurring across China. The station-level analysis provides granularity, while the provincial and national analyses offer a larger-scale perspective on temperature trends.

5.3.1 Meteorological Station Analysis

In the OLS regression analysis conducted on 139 valid meteorological stations, findings indicated that for 13 stations, the null hypothesis of no significant temperature trend could not be rejected, corresponding to the 'No Defined Trend' section in Figure 4. Conversely, the remaining 126 stations displayed significant temperature trends, with 125 stations observing a warming trend ('Positive Trend') and one station noting a cooling trend ('Negative Trend'), as illustrated in the figure. This suggests a prevalent increase in temperature across China's meteorological stations.

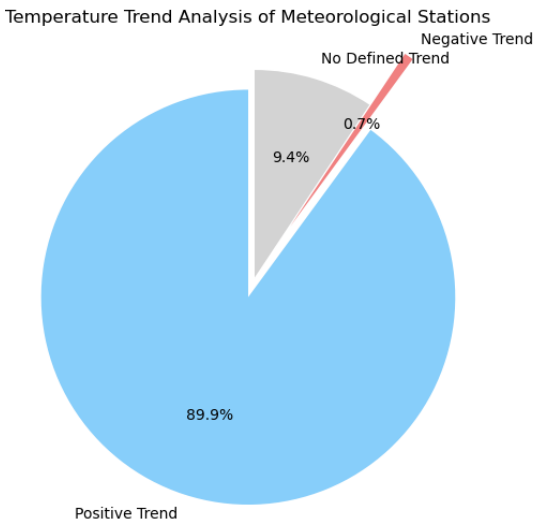

Figure 4: Distribution of temperature trends across meteorological stations, showing the majority with a warming trend, a minority with no defined trend, and a single station with a cooling trend.

5.3.2 Provincial Analysis

OLS regression analysis of temperature data from 32 provincial-level administrative regions shows that the P-Value of two provinces is higher than 0.05, indicating that there is no significant temperature trend in these regions. Among the provinces where the P-Value showed a statistically significant trend, only one showed a cooling trend, while the remaining provinces showed a warming trend. The analysis highlights major warming trends across China's provinces.

5.3.3 National Analysis

Table 2 encapsulates the statistical outcomes for various regions across China, delineating the significance and magnitude of temperature changes over a 43-year period. **The most pivotal finding of this study is the average temperature increase of 1.38°C across China over the past 43 years.**

Country	Year Coefficient	P-Value	Change (°C)
North	0.038079194	7.48×10^{-8}	1.637405
East	0.043579381	3.74×10^{-13}	1.873913
Northwest	0.035370933	4.73×10^{-10}	1.520950
Southwest	0.029438889	1.18×10^{-11}	1.265872
Northeast	0.026294584	0.000372161	1.130667
Central	0.033580785	1.63×10^{-9}	1.443974
South	0.015255106	0.003376299	0.655970
Whole China	0.032143007	1.41×10^{-10}	1.382149

Table 2: OLS Regression Results for Temperature Trends Across China and Its Regions

6 Discussion

6.1 Validating Research Correctness

The work conducted by NASA's Goddard Institute for Space Studies (GISS) has revealed a pronounced increase in the average global temperature by at least 1.1°Celsius since the year 1880. The period post-1975 has seen the majority of this warming, occurring at a rate of about 0.15 to 0.20°C per decade, with the temperature rise over land outpacing that of the oceans [5, 6, 7, 8, 9].

Specific to Mainland China, the investigation by Ren et al. [10] highlights two distinct warming phases: a moderate one between the 1930s and 1940s, and a more pronounced one from the mid-1980s onwards. The temperature anomaly charts depict an accelerated warming trend after 1980, significantly higher than the long-term average, with an estimated annual increase ranging between 0.02 to 0.03°C, as shown in Figure 5[10].

The conclusions of my study are consistent with those of the wider scientific community, and the calculated temperature rises are not too different from those recognized in the scientific literature. This consistency demonstrates that the methods I used in data collection, cleaning, and analysis were correct and that the findings are trustworthy.

6.2 Essential Learnings from Experimental Process

Appropriate Data Sources My search for weather datasets prioritized completeness, freeness, and ease of use. Since the China Meteorological Administration only provides free data sets for the last seven days, I turned to international public data sets and finally selected the Global Historical Climatology Network Daily (GHCNd) dataset. It offers free access to historical data from 100,000 meteorological stations in 180 countries, including extensive historical data from 226 stations in China. The data quality of GHCNd is highly rated online. Furthermore, GHCNd's global scope allows for an expansion of my research from China's climate analysis to future global climate trends, and enables comparisons of climate characteristics across different countries and regions [3].

Data Cleansing Strategies During the examination of downloaded datasets, it is crucial to discern data shortcomings and perform comprehensive data cleaning. This investigation focused on temperature fluctuations, necessitating consistent daily temperature records and verification of data normalcy. My approach involved graphically representing data for each site, similar to the

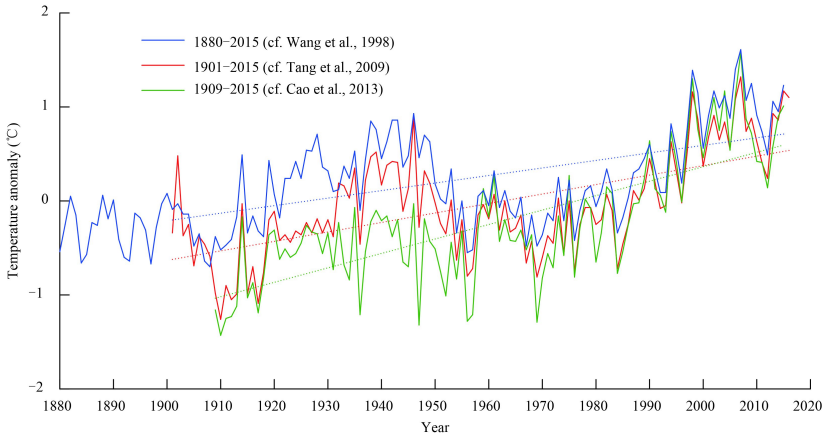

Figure 5: Illustration of the temperature increase in Mainland China post-1980, as compared to the long-term average.

depiction in Figure 6, aiding in the identification and removal of incomplete records. This process was iterative, reflecting stringent standards for data integrity.

Python as an Analytical Tool Python's extensive libraries are essential for this study's data analysis. The `pandas` library is used for data cleaning and processing [11], `statsmodels` for statistical methods [4], `matplotlib` for visualization [12], and `geopandas` for integrating data with maps [13]. This project has leveraged only the basic functionalities of these libraries, following guidelines from their official documentation. Mastery of these tools will be an area of focus for future research.

7 Conclusion & Future Work

This research utilized historical weather data and scientific methods to test the hypothesis that there was no significant change in China's temperature over the study period. It uncovered a warming trend in most regions of China from 1980 to 2022, with an average rise of 1.38°C, or an annual increase of 0.03°C. Rigorous data cleaning and OLS regression analysis were employed to calculate P-Values and reject the null hypothesis. The findings were presented both visually and numerically, offering a clear depiction of climate change.

The project aimed to deepen the understanding of global climate change complexities and provide solid, evidence-based support for Earth's conservation. It has strengthened my resolve to advocate for our planet's protection and to contribute meaningfully to environmental preservation.

This research journey has opened new avenues in data analysis for me. Moving forward, I aim to extend my study on a global scale. The objective is to discern patterns in global temperature variations, compare climatic trends across different nations, and incorporate economic data as a crucial variable. This will enable an in-depth exploration of how human activities influence climate change. The ultimate goal is to provide valuable insights and actionable data for policymakers, contributing to informed decision-making in addressing global climate challenges.

Figure 6: Visualization of daily temperature data for each meteorological station, aiding in data cleaning and integrity assessment.

8 Reflective Analysis

This project was a journey of learning and overcoming challenges. Initially, I underestimated the task, believing every step would be straightforward. However, I quickly realized that each issue demanded serious attention and effort. Mastering the process of downloading data took two days, while scripting for data processing extended over a week. There were also pleasant surprises; Python's libraries greatly simplified the statistical analysis. The entire process taught me the value of persistence.

Visualizing the results was a triumph. Making data visually intuitive not only clarified findings for the audience but also enriched my own understanding. Yet, relying on just 139 stations for nationwide analysis now seems limiting.

Given another opportunity, I'd seek more comprehensive data and explore diverse calculation methods for a robust analysis. Also, I'd love to create detailed visuals correlating temperature changes with terrain, adding another layer of insight to the findings. This experience was a profound lesson in facing challenges head-on and the power of thorough, thoughtful research.

References

[1] Jiayou Chao Jiecheng Song, Guanchao Tong. Data driven pathway analysis and forecast of global warming and sea level rise. *Scientific Reports*, 13(1):15107, 2023.

[2] IPCC. Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 2021.

[3] Global historical climatology network daily. `https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily`. Accessed: 2023-12-02.

[4] statsmodels: Statistics in python. `https://www.statsmodels.org/stable/index.html`.

[5] NASA Goddard Institute for Space Studies. Giss surface temperature analysis (gistemp). `https://data.giss.nasa.gov/gistemp`, 2023. Accessed: 2023.

[6] NASA Earth Observatory. 2022 tied for fifth warmest year on record. `https://earthobservatory.nasa.gov/images/150125/2022-tied-for-fifth-warmest-year-on-record`, 2023. Accessed: 13 January 2023.

[7] NOAA National Centers for Environmental Information. Assessing the global climate in 2021. `https://www.ncei.noaa.gov/news/global-climate-202112`, 2022. Accessed: 10 January 2022.

[8] NASA Earth Observatory. Why so many global temperature records? `https://earthobservatory.nasa.gov/features/GlobalWarming/page3.php`, 2015. Accessed: 21 January 2015.

[9] NASA Earth Observatory. Global warming. `https://earthobservatory.nasa.gov/features/GlobalWarming`, 2010. Accessed: 3 June 2010.

[10] Guoyu Ren, Yihui Ding, and Guoli Tang. An overview of mainland china temperature change research. *Journal of Meteorological Research*, 31(1):3–16, 2017.

[11] pandas: powerful python data analysis toolkit. `https://pandas.pydata.org/`.

[12] Matplotlib: Visualization with python. `https://matplotlib.org/`.

[13] Geopandas: Python tools for geographic data. `https://geopandas.org/`.