

[**https://github.com/lhansen77/APPLIED-DATA-SCIENCE-FOR-UNDERSTANDING-GLOBAL-WARMING-TRENDS**](https://github.com/lhansen77/APPLIED-DATA-SCIENCE-FOR-UNDERSTANDING-GLOBAL-WARMING-TRENDS)

**Project 1: Proposal & Data Selection Group Assignment (Milestone 1)**

**DSC450-Applied Data Science**

**Ryan Weeks, Sarah Yawn, Lisa Hansen**

**09/27/2025**

**Abstract**

This project explores long-term global temperature trends using a cleaned and feature-engineered dataset spanning multiple centuries. The analysis focuses on identifying seasonal patterns, regional anomalies, and decade-level shifts in climate behavior. Methods include time-series visualization, seasonal decomposition, and geographic mapping, supported by Python-based tools in a Jupyter environment. Anticipated challenges include sparse historical data, regional gaps in coverage, and maintaining visual clarity across time scales. The research aims to answer key questions about how temperature trends vary by region and season, and whether anomalies align with known climate events. By combining statistical rigor with accessible visual storytelling, this project contributes to a deeper understanding of global climate change and its uneven impact across geographic regions.

**The Domain**

The dataset used in this project falls within the domain of climate science and environmental analytics. Specifically, it focuses on global temperature records collected over multiple centuries, enabling the exploration of long-term climate change patterns, seasonal variation, and regional anomalies. This domain intersects public health, policy planning, and sustainability efforts, making the analysis relevant to both scientific inquiry and community impact.

**References**

Lindsey, R., & Dahlman, L. (2025, May 29). Climate change: Global temperature. *NOAA Climate.gov*. <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>  
NASA. (n.d.). Evidence. *NASA Science*. <https://science.nasa.gov/climate-change/evidence/>  
Intergovernmental Panel on Climate Change. (2023). *Climate change 2023: The physical science basis*. <https://www.ipcc.ch/report/ar6/wg1/>  
Jones, P. D., & Moberg, A. (2003). Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of Climate, 16*(2), 206–223. <https://doi.org/10.1175/1520-0442(2003)016><0206:HALSSA>2.0.CO;2  
Rahmstorf, S., & Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences, 108*(44), 17905–17909. <https://doi.org/10.1073/pnas.1101766108>  
Hansen, J., Sato, M., & Ruedy, R. (2012). Perception of climate change. *Proceedings of the National Academy of Sciences, 109*(37), E2415–E2423. <https://doi.org/10.1073/pnas.1205276109>  
Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research, 47*(1–2), 123–138. <https://doi.org/10.3354/cr00953>  
Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society, 90*(8), 1095–1107. <https://doi.org/10.1175/2009BAMS2607.1>  
Lindsey, R., & Dahlman, L. (2024, January 18). Climate change: Global temperature. *U.S. Department of Energy*. <https://www.energy.gov/sites/default/files/2024-02/093.%20Rebecca%20Lindsey%20and%20Luann%20Dahlman%2C%20NOAA%2C%20Climate%20Change_%20Global%20Temperature.pdf>  
National Oceanic and Atmospheric Administration. (2024). *2024 Global climate report*. <https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202413>  
Hansen, L. (2025). *Ocean climate change: Historical trends and regional anomalies* [Unpublished student report]. GitHub. <https://github.com/lhansen77/LisaHansen-Portfolio/blob/main/project2.md>

**The Data**

The primary dataset used in this project is *GlobalTemp\_Cleaned\_CountryContinent.csv*, which contains monthly temperature records by country from 1743 to 2013. It includes added features such as Continent and UN\_Region for geographic aggregation. A trimmed version, *GlobalTemp\_Post1850.csv*, restricts analysis to post-1850 data to reduce noise from sparse early records. A third dataset, *Annual\_Anomalies.csv*, contains annual temperature anomalies relative to the 1951–1980 baseline.

The data was cleaned using Python and pandas in a Jupyter environment. Missing values were removed, and features such as Season, Decade, and Year were engineered to support time-based analysis. Aggregated views by country and time period allow for comparative analysis across continents and decades. The format supports statistical modeling and visual storytelling, enabling the identification of patterns, detection of outliers, and communication of climate insights.

**Research Questions? Benefits? Why Analyze the Data?**

This project investigates the following research questions:

* How have global temperatures changed across regions?
* Which continents show the strongest long-term warming?
* Do equatorial regions remain more stable than continental climates?
* Are seasonal patterns consistent across continents?
* How does dataset selection (full vs. post-1900) influence conclusions?

These questions are addressed through rolling averages, anomaly detection, and regression modeling. The benefits include identifying uneven warming rates, forecasting future anomalies, and providing a foundation for visual storytelling that spans past, present, and future. The analysis supports climate science, public awareness, and policy planning by offering statistically grounded insights into regional climate shifts.

**The Method**

The project uses a combination of exploratory data analysis (EDA), statistical modeling, and forecasting. Key methods include:

* **EDA**: Sanity checks, seasonal pattern analysis, anomaly detection, and regional comparisons.
* **Regression Modeling**: Estimating warming rates using Anomaly ~ Year + Continent (+ Season), with interaction terms to test for differences in warming rates.
* **Forecasting**: Applying ARIMA and Prophet models to anomaly series by continent. Prophet generally captured upward trends more reliably, while ARIMA occasionally underfit trajectories.
* **Scenario Comparison**: Evaluating how early, noisier data alters regression slopes and forecast outcomes.

These methods provide statistical evidence of uneven warming, model-based projections to 2050 or 2100, and comparative insights into dataset reliability.

**Potential Issues**

Anticipated challenges include:

* Sparse and inconsistent early records (pre-1850), which introduce volatility.
* Regional gaps in data coverage, especially in Oceania and South America.
* Model limitations in capturing short-term volatility, as seen in Prophet forecasts that underestimate recent anomaly spikes.
* Visual clarity across time scales, requiring careful smoothing and aggregation.
* Time constraints in refining models and interpreting interaction effects.

These issues may affect the schedule and interpretation, but are mitigated through dataset trimming, rolling averages, and incremental analysis.

**Concluding Remarks**  
This project offers a structured, reproducible approach to analyzing global temperature trends across centuries. By combining cleaned datasets, statistical modeling, and visual storytelling, it reveals uneven warming rates, seasonal variability, and the accelerating pace of climate change. The findings highlight the importance of dataset selection, model choice, and regional context in climate analysis. Ultimately, the work contributes to a deeper understanding of global climate dynamics and supports informed decision-making in science and policy.