Proposal

al4771, cx2347, jh4977, km4071, xf2290

2025-11-05

Group Numbers

Jingtong He(jh4977), Chuyuan Xu(cx2347), Anqi Li(al4771), Xiwei Fu(xf2290), Kaicheng Mo(km4071)

Tentative Project Title

ClimaScope: Forecasting the Empire Skies

Motivation

Climate variability and extreme weather increasingly affect New York State. Understanding local patterns is vital for public health, agriculture, and infrastructure. Using five years of weather data, this project visualizes and models temperature, precipitation, and wind speed to reveal key climatic drivers and support regional climate planning.

Intended Final Products

Interactive Visualizations showing spatial and temporal patterns of temperature, precipitation, and wind speed across New York counties from 2016–2020.
Cleaned and Integrated Dataset with detailed data dictionary and quality summary.
Modeling and Forecasting Results including regression analyses, mixed-effects models, and short-horizon time-series forecasts with evaluation metrics.
Final Report and GitHub Repository containing reproducible code, documentation, and all figures for transparent analysis.

Anticipated Data Sources

The meteorology datasets were downloaded from [**"Daily meteorological Gridmet variables by United States administrative boundaries", Harvard Dataverse**](https://doi.org/10.7910/DVN/FYTME3). To monitor the changing patterns in the past seasons and years, meteorological data from 2016 to 2020 on the county level.

Corresponding county information was downloaded from [**tiger/line Shapefiles**](https://www.census.gov/cgi-bin/geo/shapefiles/index.php). Due to the lapse in federal funding, portions of the [**United States Census Bureau**](https://www.census.gov/) were unavailable. `tl_2020_state` and `tl_2024_county` were downloaded to obtain additional New York state and county information to facilitate further visualizations and analysis.

Planned Analyses/ Visualizations

Exploratory Data Analysis: Visualize spatial and temporal patterns of temperature, precipitation, and wind speed using histograms, density plots, heatmaps, and county-level maps. Examine correlations and temporal dependencies among key variables.

Spatiotemporal Association Modeling: Fit regression and mixed-effects models to assess associations between meteorological outcomes and geographic factors, using two-part models for precipitation where needed. Present results with coefficient plots and diagnostics.

Visualization Deliverables: Deliver interactive and static visualizations, including maps, seasonal plots, and correlation displays, summarizing key findings clearly and reproducibly.

Coding Challenges

Key challenges include integrating large meteorological datasets, managing spatial data for county-level mapping, and aligning geographic identifiers. Implementing regression and mixed-effects models efficiently in R and ensuring reproducibility through clear, well-documented code will be essential.

Planned Timeline

Nov 1–5: Define project scope, motivation, and research questions.
Nov 6–15: Collect, clean, and merge meteorological data into an analysis-ready dataset.
Nov 16–25: Explore spatial and temporal climate patterns through descriptive statistics and visualizations.
Nov 26–Dec 5: Fit regression and mixed-effects models to identify key climatic associations.
Dec 6–15: Summarize findings, visualize results, and compile the final report and GitHub documentation.

Project Structure

Introduction : Present the motivation and significance of studying climate variability across New York State.
Data Cleaning : Compile and standardize five years of meteorological data, addressing missing values and ensuring consistency across counties.
Exploratory Analysis and Visualization: Examine spatial and temporal trends in temperature, precipitation, and wind speed using descriptive plots and maps.
Statistical and Predictive Modeling: Apply regression and mixed-effects models to assess associations between meteorological and geographic factors.
Results and Discussion: Summarize major findings and their implications for climate monitoring, infrastructure, and regional resilience.