## Non-Parametric Density Estimation (PART 5)

(PART 5: Andrew)

## Introduction

## Part C of the Risk-Profile Analysis section of the group assignment focuses on using Non-Parametric Density Estimation to analyze historical hurricane data. Based on past hurricane trajectories and their severity, the analysis targets areas with high location risks. This approach provides valuable insights for disaster preparedness and regional planning, particularly in areas prone to severe weather events. The links provided in the Data Collection and Preparation section of the assignment were beneficial in obtaining the necessary TroPyCal Python library API information to integrate the dataset into the code. The HURDAT database from the National Hurricane Center offered a variety of data on hurricanes and cities that have been heavily impacted by them over the past 25 years. This database is a crucial resource as it contains extensive historical records, including detailed information on storm paths, intensities, and frequencies, which are very important for assessing risk patterns.

## Setup

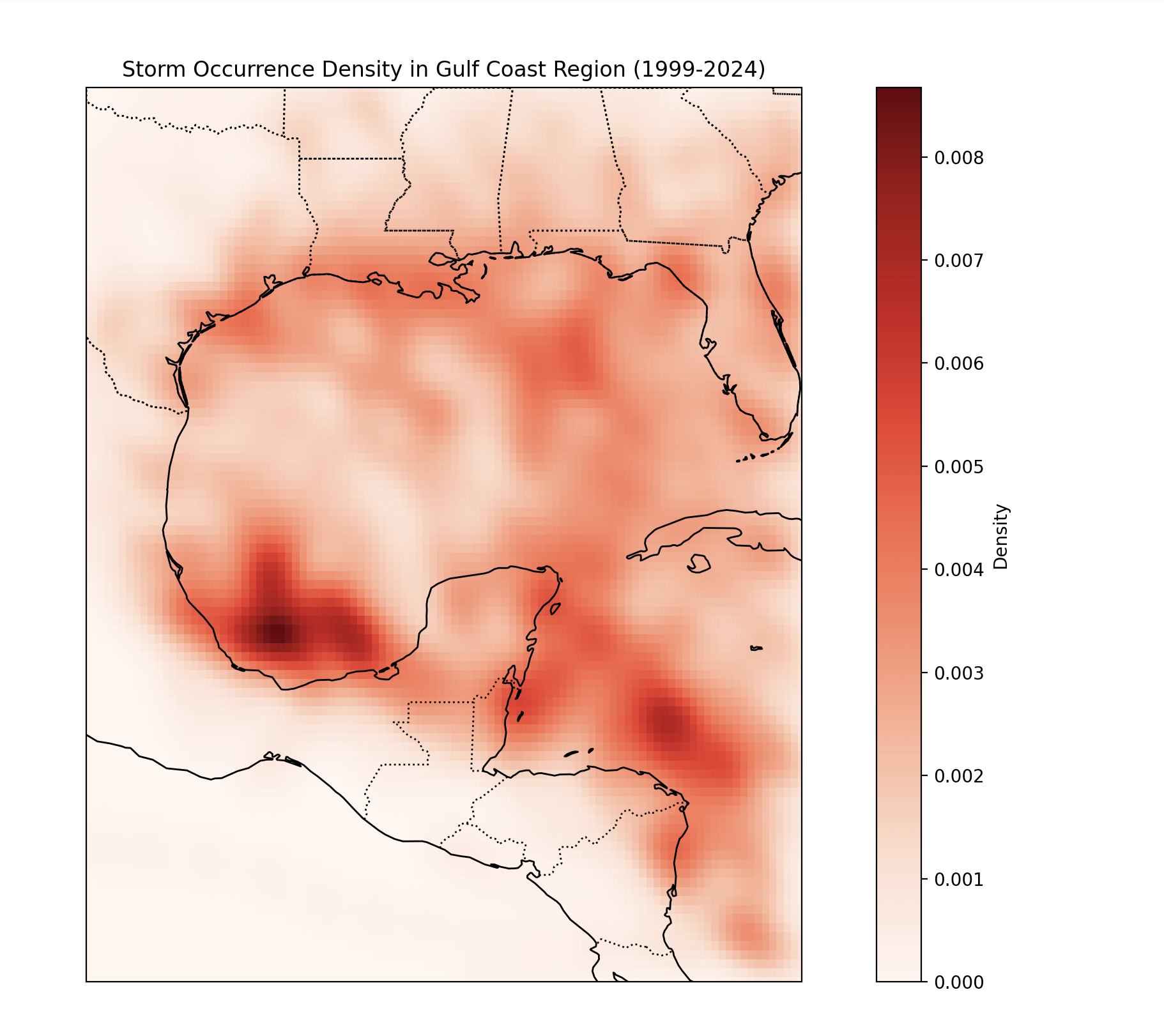
While gathering and organizing the required information was initially challenging due to the dataset's size and complexity, collaboration within the group helped resolve these issues efficiently. This teamwork also facilitated preprocessing the data and saving it as a .pkl (pickle) file for streamlined loading and analysis. Key fields such as latitude, longitude, and time were standardized, with time converted to datetime format to enable precise year-based filtering. For the analysis, the Gulf Coast region was defined with specific latitude and longitude bounds: latitude between 10° and 35° N and longitude between -100° and -80° W. This geographic focus makes sure that the dataset captured storms that had the most direct impact on Gulf Coast areas. Having a large sample size spanning 25 years was very valuable for producing reliable and detailed predictions about location risks. The extensive dataset allowed for the identification of clear trends and patterns in storm trajectories and intensities over time. A custom filtering function, *get\_gulf\_storms\_data*, was implemented to isolate storms that either formed within or entered the Gulf Coast region during this timeframe.

Kernel Density Estimation (KDE)

Following data filtering, Kernel Density Estimation (KDE) was applied to estimate storm occurrence density. The KDE, a non-parametric statistical technique, was particularly beneficial for this analysis as it does not assume a fixed data distribution. This flexibility is ideal for hurricane data, as storm paths and intensities are unpredictable and often influenced by numerous environmental factors. By estimating the probability density function of storm occurrences, KDE offered a clear visual representation of high-risk areas. The bandwidth parameter of the KDE was adjusted carefully to balance smoothness and clarity. This adjustment ensured that the density map effectively highlighted high-density areas without over-smoothing, which could obscure important patterns or dilute the visual. The density map visualization was created using Cartopy, a Python library for geospatial data visualization. The map incorporated geographic context, including coastlines, landforms, and state boundaries, to provide a comprehensive visual reference. The KDE output was overlaid as a color gradient, where warmer colors indicated areas of higher storm density, while lighter shades indicated regions with less frequent storm activity. This overlay made it possible to pinpoint high-risk areas with high amounts of clarity, making sure the visualization was both informative and accessible.

Results and Analysis

Along the Gulf Coast, the lower half of Mexico emerged as a significant hotspot for storm activity, particularly along its coastal areas. These regions consistently experienced high storm frequencies, highlighting their vulnerability to severe weather. The map showed dense clusters of activity that correlated with known hurricane landfall zones, underscoring the recurring nature of these events. Similarly, the coastal stretch near Honduras and the region between Cuba and the Yucatán Peninsula exhibited substantial storm activity, which is consistent with storm systems originating in the Caribbean. This area acts as a corridor for hurricanes, channeling storms into the Gulf of Mexico, where they often intensify before making landfall. In the United States, states such as Texas, Louisiana, Mississippi, Alabama, and Florida showed moderate but persistent storm densities. These states are frequently impacted by hurricanes, particularly those originating in the Gulf. While the density in these areas was not as high as in regions further south, the consistent storm activity still poses a considerable risk.



Despite its strengths, KDE has limitations. The smoothing process, while effective for general patterns, can obscure specific trajectory details or rare, intense storms, potentially underrepresenting critical outliers. For example, highly destructive hurricanes like Katrina or Harvey, which had significant impacts, may not be adequately highlighted in a smoothed density map. A potential improvement would involve integrating additional statistical techniques, such as clustering or outlier detection, to better identify and emphasize these extreme events. Additionally, expanding the dataset to cover a longer timeframe—such as 50 years instead of 25—could provide a more comprehensive understanding of long-term hurricane trends and their evolving impacts, particularly in the context of climate change.