Interactive Information Visualization Project Report

Team name : **Great Studies**

Team members :

Rayane Dahasse (rayane.dahasse@gmail.com)

Ray Cheng (raycheng.cs11@nycu.edu.tw)

Rabii Lamharzi Alaoui (rabiialaoui10@gmail.com)

Teachers :

Dr. Petra Isenberg (petra.isenberg@inria.fr)

Pr. Anastasia Bezerianos (anastasia.bezerianos@universite-paris-saclay.fr)

Pr. Aikaterini Batziakoudi (aikaterini.Batziakoudi@inria.fr)

2024 – 2025

Summary

Introduction

The goal of this project is to explore how weather conditions impact students' daily lives, academic performance, and engagement with campus resources and activities. Weather can influence study habits, class attendance, mental well-being, and participation in extracurriculars, ultimately shaping the overall student experience.
By understanding these effects, universities can develop strategies to support student success, promote equity, and enhance resilience in the face of varying and potentially extreme weather conditions. This research will provide insights that can inform campus policies, resource allocation, and student support systems to ensure a more adaptable and inclusive learning environment.

In order to address this topic, we started by looking for datasets related to our project, we then started discussing them and chose to keep 2 of them and combine them.

Technical Realisation

Our dataset is a combination of the following datasets

School Student Daily Attendance (2018-2019)	● This dataset, sourced from Kaggle and provided by the NYC Department of Education, tracks daily student attendance in New York City schools from September 5, 2018, to June 26, 2019.
Historical Weather Data for NYC (2018-2019)	● This dataset, sourced from Open-Meteo and based on ECMWF IFS reanalysis dataset, provides historical weather conditions for NYC during the same period (2018-2019).

In order to analyze the impact of weather on school attendance, we have implemented the following data processing and aggregation steps

1. **Date Formatting and Merging:** The date format in both datasets was standardized to YYYY-MM-DD. The datasets were then merged by matching dates.

2. **Sorting:** The combined dataset was sorted first by 'School DBN' and then by 'Date' within each 'School DBN'.

3. **Weather Code Mapping:** Weather codes in the weather dataset were mapped to their corresponding WMO code descriptions. For example, 2 as "State of sky on the whole unchanged", 5 as "Haze"

4. **Column Removal:** Unnecessary columns, including maximum and minimum temperature, wind speed, snowfall, and rainfall, were removed.

5. **Aggregation:** For each 'School DBN' and 'Date' combination(Data in same date were combine together, calculated the sum of the enrollment, absent, and present) , the following aggregations were calculated:

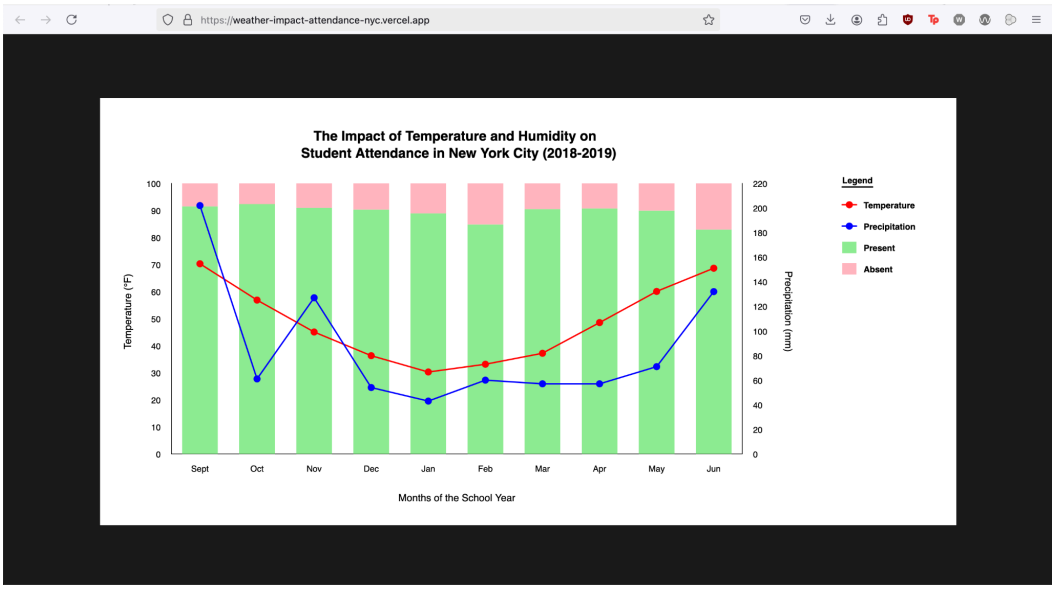
    - Mean of 'absent' students.
    - Mean of 'present' students.
    - Mean of 'absent' rate.
    - Mean of 'present' rate.
    - Temperature stays the same
    - Precipitation stays the same
    - The weather code descriptions were retained.

6. The final dataset includes the following columns: Date, Absent_mean_NY, Absent_rate_mean, Present_mean_NY, Present_rate_mean, temperature_mean (°C), precipitation (mm), and description (weather description)

7. As we were shown in class, we capitalized on using the p5.js library to create our visualization. Here are the main steps we undertook:

    - Created preload() function to load CSV data using loadTable()
    - Created setup() function to:
        i. Set canvas size
        ii. Get data columns from CSV
    - Created calculateMonthlyAverages() function to:
        i. Process temperature (averages)
        ii. Process precipitation (monthly totals)
        iii. Process attendance data
    - Created draw() function with:
        i. Drawing axes and labels
        ii. Creating stacked bar chart for attendance
        iii. Drawing temperature line (red)
        iv. Drawing precipitation line (blue)
        v. Adding legend and title
        vi. Centering everything on canvas

Here is what the initial version of our data visualization looks like:



The data visualization is also available online here:

https://weather-impact-attendance-nyc.vercel.app/

You can access the GitHub source code for our project here:

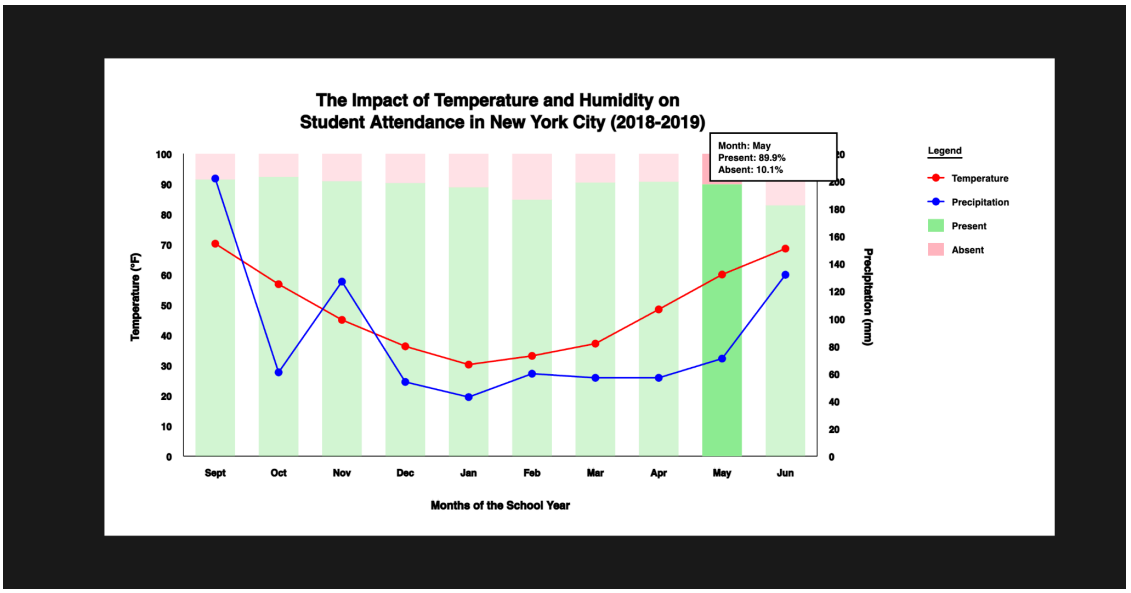
https://github.com/Ray-Cheng-11/Weather_Impact_Attendance

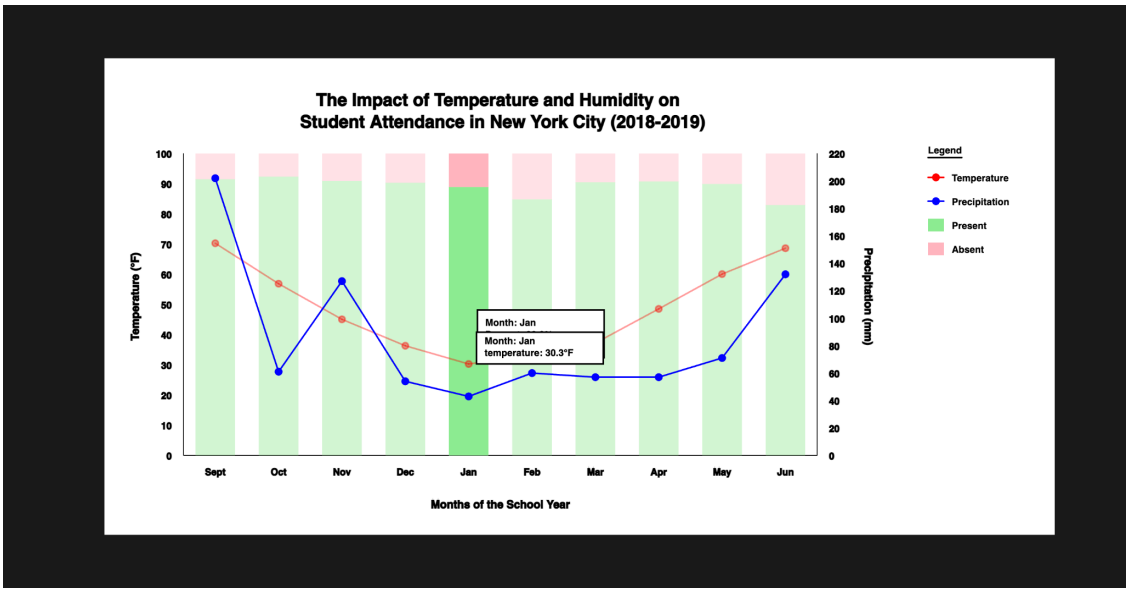
Feature of the tool

After a productive team brainstorming session, we decided to enhance our sketch by implementing a hovering feature that provides users with additional information about specific months. This feature improves interactivity by displaying detailed data when the user hovers over a temperature or precipitation point with their mouse.

To achieve this, we continuously check whether the mouse coordinates on the canvas align with any data points. If a match is detected, we update the hovering state and trigger the drawInfo function, which presents relevant details in real time. This allows users to gain deeper insights into weather trends without needing to click or navigate away.

Below is an example of how the sketch appears with the hovering feature in action.





Interesting Findings

By analyzing the dataset using our tool, we have identified two significant peaks in the student absence rate in New York. The first peak occurs in February, which aligns with the lowest temperatures of the year. This correlation suggests that cold weather conditions, potentially accompanied by snow and icy roads, may contribute to increased student absences due to transportation challenges, illness, or discomfort.

The second peak is observed in June, where we notice a combination of extremely high temperatures and a significant increase in precipitation. This likely indicates elevated humidity levels, which could make the weather feel even more uncomfortable. The combination of heat and humidity may lead to physical discomfort, fatigue, or even weather-related health concerns, potentially explaining the rise in absences during this period.

These patterns provide valuable insights into how weather conditions influence student attendance, helping us better understand seasonal trends and their impact on academic participation.

Improvement and Limitations

One possible improvement for our project is increasing data granularity by analyzing daily or weekly trends instead of monthly averages, which would provide a more detailed understanding of short-term weather effects on student attendance. Another enhancement could involve integrating additional weather factors, such as wind speed and air quality, to offer a more comprehensive analysis of environmental influences on student behavior.

However, our tool has some limitations. One major challenge is distinguishing correlation from causation, as other factors like school policies or seasonal illnesses may also contribute to absence rates. Additionally, the accuracy of our findings depends on the completeness and reliability of the dataset, meaning that missing or inconsistent data points could affect the validity of our conclusions.