Milestone 1

Data Visualization COM-480
Spring Semester 2025

Group Members:

Antonio Mari
SCIPER: 377119

Marco Scialanga
SCIPER: 369469

Ahmed Abdelmalek
SCIPER: 344471

Instructor: Laurent Vuillon

1. Problematic

In many large cities, personal safety is a concern for everyone. People who are new to a city are often unfamiliar with its neighborhoods and streets, so they rely on navigation apps to get around. A route that appears straightforward on Google Maps might lead travelers through high-risk areas, exposing them to theft, assault, or other dangers. Exchange students and tourists are particularly vulnerable to street crime, and they would benefit from knowing which areas to avoid while visiting.

Our goal is to develop a visualization-driven tool that predicts and highlights potential safety risks across various parts of the city. We aim to demonstrate our idea in a real-world setting by making the tool work for the city of Chicago. Since Chicago is home to more than 20 universities, many students could potentially benefit from this work. By analyzing historical crime data, we aim to give users a clear picture of which places might carry greater risk and why.

Features we plan to implement:

- After selecting the desired time, starting point, and destination, the tool analyzes paths proposed by common navigation apps and computes a "safety score" for each route. It then suggests the best option by balancing distance and safety.
- Once a route is chosen, the tool will display useful statistics, charts, and safety advice about the risky "hot spots" along the way.

The purpose of **SafetyFirst** is to use intuitive maps and interactive charts to help users make informed travel decisions, increasing awareness of localized crime trends. Ultimately, this project aims to foster a broader understanding of the factors that contribute to urban safety.

2. Dataset

For this project, we will use the Crimes - 2001 to Present dataset from the City of Chicago's Data Portal. This dataset contains over 8 million crime records, with key attributes such as crime type, date and time of occurrence, location (including latitude and longitude), and whether an arrest was made. The data is updated frequently, ensuring a reliable and up-to-date analysis of crime patterns in the city.

We assessed the dataset's quality by analyzing its completeness and consistency. The 2025 subset contains 38,516 records, with crime reports spanning different locations across the city. Most records contain valid timestamps and geographic coordinates, making it a perfectly suitable dataset for our analysis.

Since the dataset contains many unnecessary fields, we will refine our selection to focus on crime descriptions, timestamps, and geolocation coordinates. This preprocessing step

involves filtering out records that lack latitude and longitude data and excluding crimes irrelevant to street safety, such as financial fraud or offenses occurring in private properties. By doing so, we ensure that our analysis remains focused on public safety in outdoor environments.

Regarding interactive visualization, we will use the Leaflet JavaScript library, which integrates OpenStreetMap data. This choice ensures flexibility, ease of use, and high-quality map rendering. Leaflet allows users to interact dynamically with crime heatmaps and route safety suggestions, making it an ideal tool for our project.

3. Exploratory Data Analysis

To check the quality of the dataset, we performed an exploratory analysis to identify crime trend and patterns (note: we included 2025 data since recent information is very relevant for our project, causing the graphs below to have a sharp drop at the end, since data for this year is only until March). The number of crimes reported yearly (Fig. 1) significantly declined since the early 2000s, then remained stable from 2014, with a drop in 2020-2021 (correlated to the Covid-19).

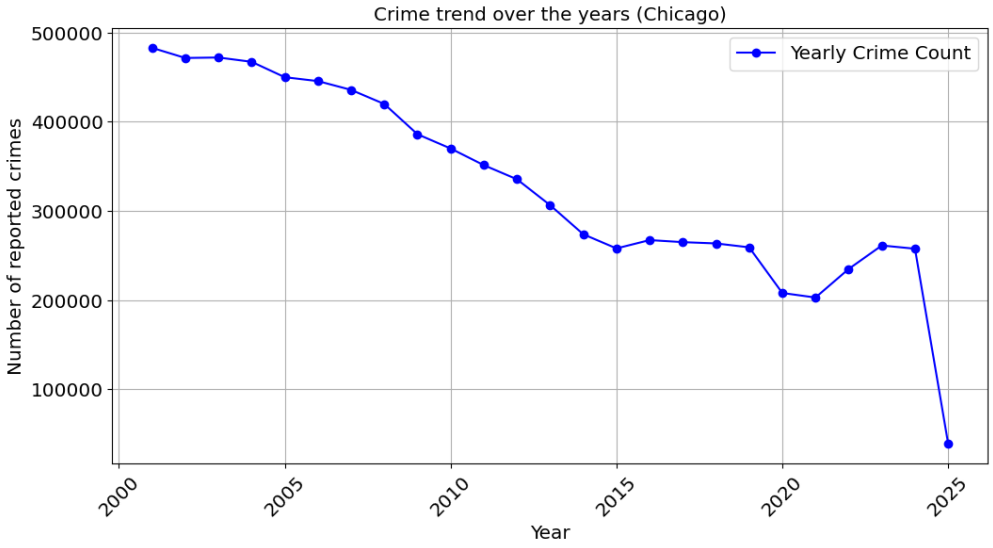

Figure 1: Number of reported crimes in Chicago year by year (2001 to present)

Among the most frequent crimes (Fig. 2), theft, battery, burglary and robbery concern anyone moving around the city. A huge proportion of crimes can be grouped into 4 broader categories (see Table 1): Assault, Theft, Sex Offense and Minor. We excluded from our analysis all crimes unrelated to public safety ("Non street Crime" in the table).

Category	Description	Examples	Num of crimes
Physical Assault	Violent crimes involving physical force	Battery, Robbery, Homicide	2 497 611
Theft	Theft or vandalism	Vehicle theft	3 085 635
Sex Offense	Sex crimes	Sexual Assault, Stalking	79 374
Minor	Crimes not harming passers-by	Prostitution, Obscenity	1 322 682
Non Street Crime	Any crime unrelated to public safety	Fraud, License violation	1 146 785

Table 1: Macro-categories for Crimes.

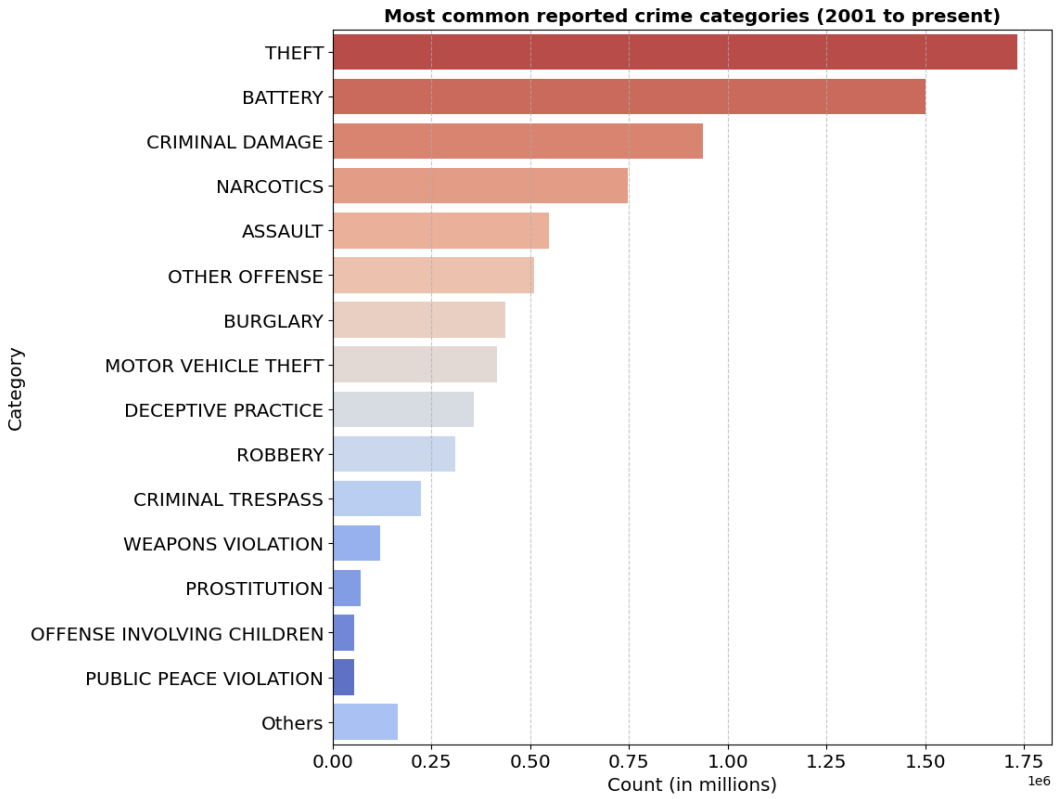

Figure 2: Most common reported crime categories in Chicago (2001 to Present)

Similarly, we grouped crimes by their **location description**:

- *Street*: crimes reported in public spaces, e.g. streets, alleys, gas stations, ATMs.
- *Public transports*: crimes targeting public transport facilities and vehicles.
- *Airport*: crimes in the airport, we ignored them because they are not helpful to suggest routes.
- *Stores*: crimes reported in many type of stores/shops/food places.
- *Non concerning*: crimes reported in private places, unreachable to city travelers.

We report the distribution of crime locations in Fig. 3. Street crimes are most numerous, so our goal is achievable with our data.

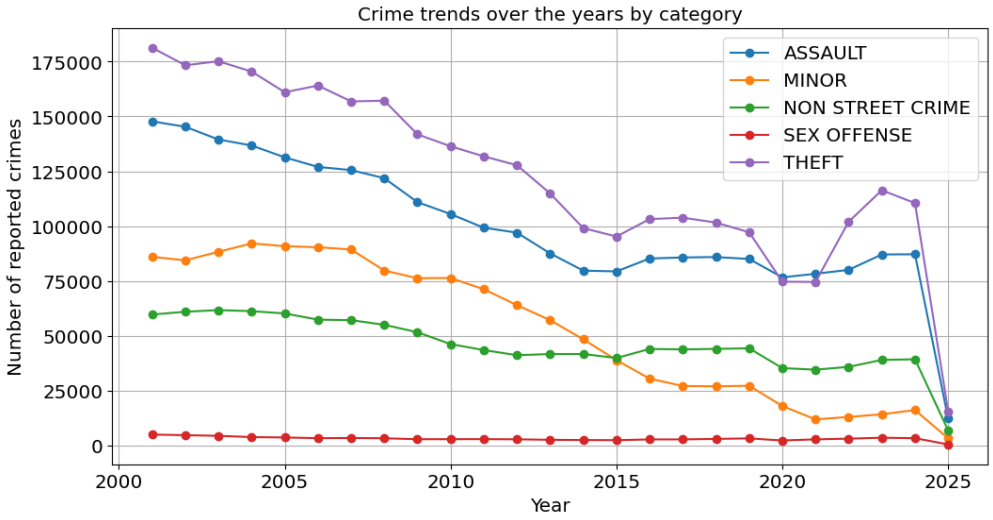

Figure 3: Crime trends over the years by crime category (2001 to present)

Finally, we generated a detailed Chicago crimes heat-map. In Fig. 5, we zoomed in on "Sex Offense" and "Assault" crimes reported in a small area of Chicago between 00:00 and 03:00.

We can easily identify "hot-spots", locations in which many crimes have been reported. The streets topology is optimal for our application, with squared tiles allowing for multiple paths with the same walking distance.

Our analysis confirmed that this dataset is ideal for our project, providing insightful crime trends that may enhance real-time risk-aware navigation.

4. Related Work

While many existing works have explored Chicago crime data, most solutions focus on providing basic visualizations of crime patterns to highlight high-risk areas. Mainly, the Chicago Data Portal provides a dashboard with heat maps, bar charts and line plots describing crime incidents in all of Chicago's police districts. In addition, users can interact with a map showing all recorded crimes starting from 2001, with the possibility of zooming in at the street level. Furthermore, on the popular data science platform Kaggle, there are hundreds of notebooks working on this dataset. Unlike the Chicago Crime Data Portal, some works here show visualizations that identify trends in time rather than just in space. Some interesting approaches include looking at the effects of lockdown on crime and at correlations between taxi tips and dangerousness of neighborhood. Many users analyze this dataset as an exercise to practice SQL and Pandas.

Our approach goes beyond static crime visualizations and instead offers a dynamic route safety analysis. While existing works provide crime maps, statistics, and temporal trends, they lack the ability to guide users through the safest paths based on historical crime data.

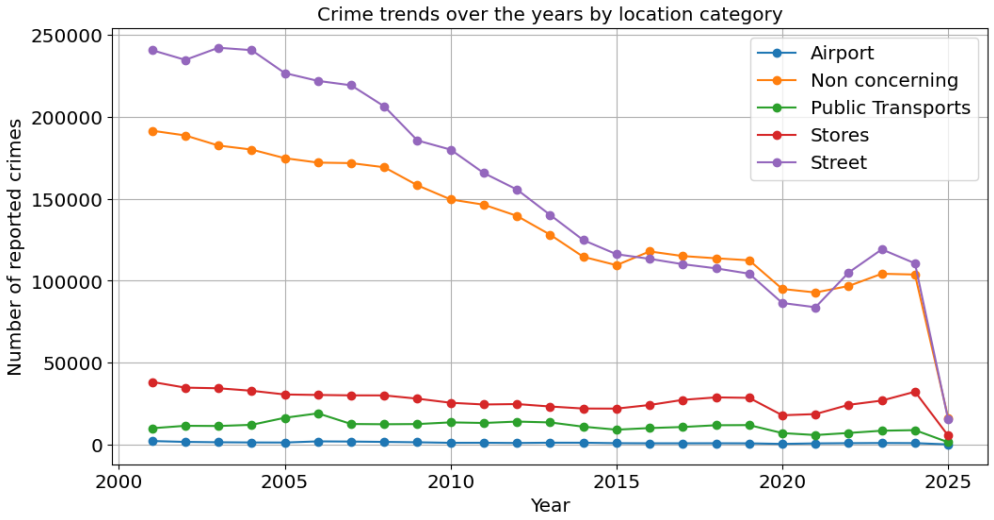

Figure 4: Crime trends over the years by crime category (2001 to present)

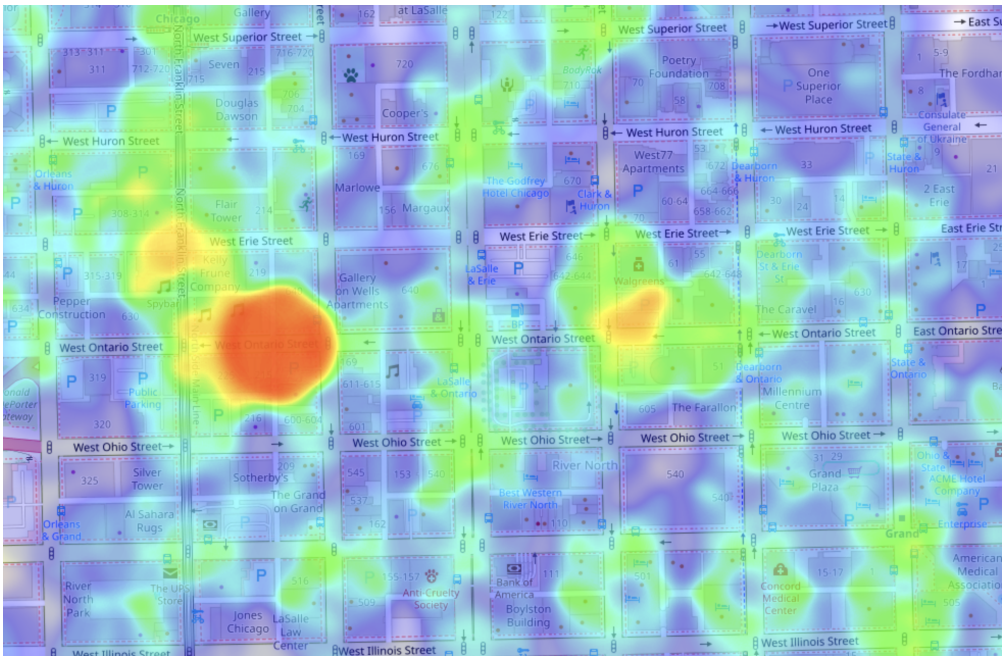

Figure 5: Heatmap of "Sex Offense" and "Physical Assault" crimes between 00:00 and 03:00 in a small area of Chicago.

By transforming raw crime data into actionable insights for everyday navigation, our project offers a unique, practical, and useful solution to enhance personal safety in unfamiliar urban environments.

In fact, we know by personal experience that a lot of EPFL students might want to study abroad for a semester. However, not all university campuses around the world are as safe as ours. Thus, after finding a large and ready to use crime dataset such as the one from

the Chicago Police, and listening to the Professor's advice on using this data to identify safe paths around a city, we decided that building a tool such as the one we described above could potentially be a valuable resource for students and an interesting challenge for this course's project.