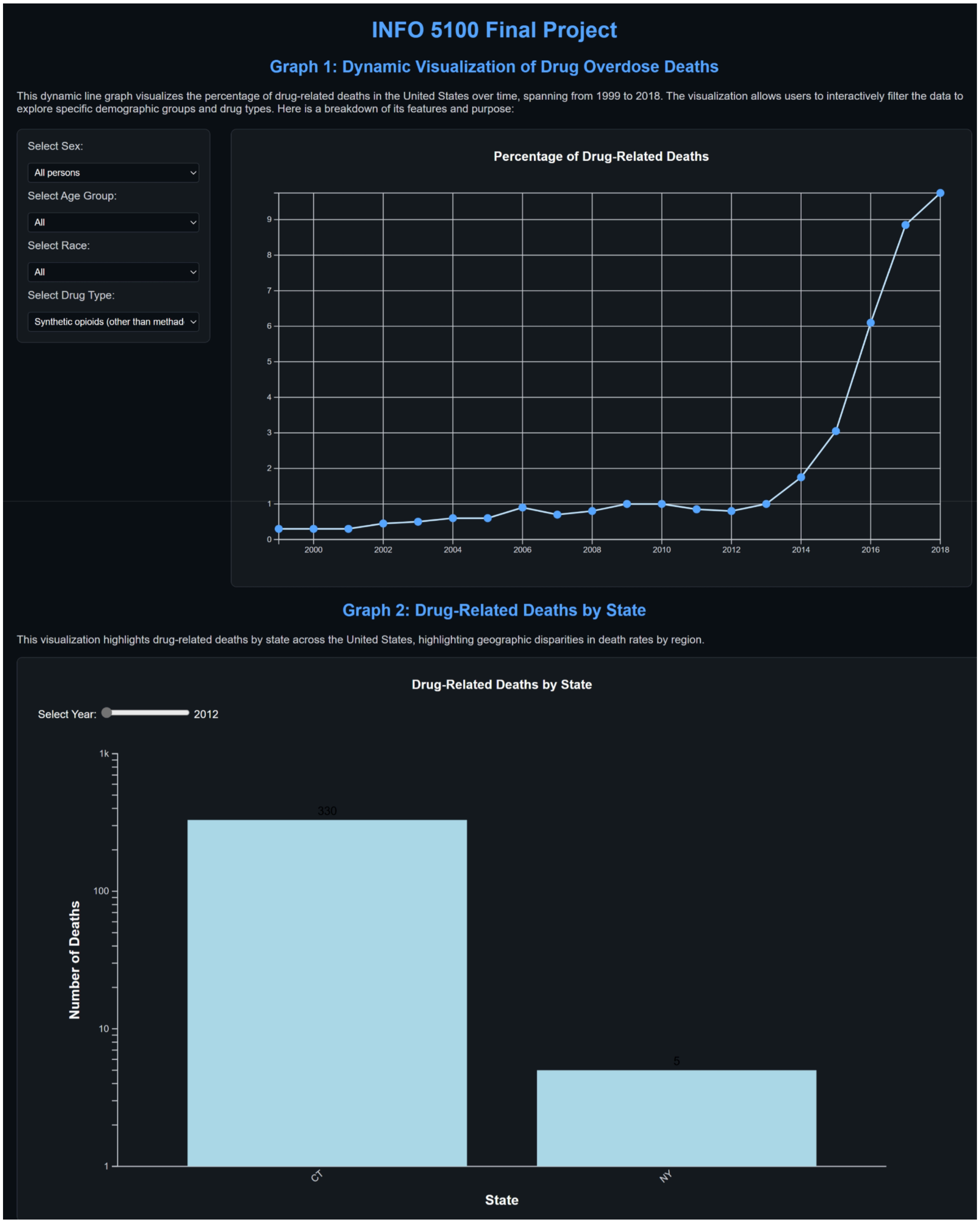
Project Report

Haneul Song, Lily Huang, Peipei Duan, Katie Zhou

Visual Screenshot



Part A: Data Description

The first dataset (data.csv) contains detailed information on incidents involving injuries and deaths, with a focus on drug-related cases. It consists of 9,202 rows and 48 columns, covering various aspects such as demographic details, locations, causes of death, and substances involved. Demographic variables include Age, Sex, Race, and Ethnicity, while location-related fields include Residence City, Residence County, and Injury City. The dataset also provides geospatial information through columns like ResidenceCityGeo, which includes coordinates for mapping purposes. Additionally, it contains information on drug involvement, listing substances such as Heroin, Cocaine, Fentanyl, and others. Supplementary fields describe injury and death details, such as the Cause of Death and Description of Injury, providing a comprehensive view of each case.

The second dataset (first_data.csv) focuses on trends related to drug overdoses, with 6,228 rows and 9 columns. Key variables include drug_type, case, and details, alongside yearly data (year) and percentages (percentage). Demographic information is also included through fields like sex, age, race, and Hispanic. This dataset offers a summarized view of drug overdose trends over time, broken down by specific demographic and case characteristics.

To prepare these datasets for analysis, First, subsets of relevant columns were selected to focus on specific research questions, such as trends involving particular drugs or demographic patterns. Missing data will need to be addressed, with strategies such as imputation or removal depending on the context. If the datasets are to be combined, common fields like year or demographic variables can serve as points of integration. Geospatial data in the first dataset may need formatting for mapping purposes, while percentage trends in the second dataset could be normalized or aggregated for better comparative analysis. These steps ensure the data is clean, structured, and ready for insightful analysis.

Part B: Design Rationale

The dynamic line graph visualizing trends in drug overdose deaths from 1999 to 2018 was designed with clarity and interpretability in mind. The primary visual channel used is position, with the x-axis representing time (years) and the y-axis showing the percentage of drug-related deaths. Continuous line marks were chosen to connect data points for each demographic group or drug type, as lines are effective in conveying trends over time. Color is employed as a secondary visual channel to distinguish between different groups or categories, making it easy to differentiate between multiple lines. Subtle gridlines were added to assist in value estimation while avoiding visual clutter. To enhance usability, filters allow users to focus on specific subsets of the data, such as sex, age group, race, or drug type, reducing overlap and clutter. The y-axis dynamically scales based on the selected filters to provide optimal resolution for the data being explored. Interactive elements, such as hover-based tooltips, provide precise information about individual data points without overwhelming the visualization. However, trade-offs include potential clutter when too many filters are selected at once, which was mitigated through selective filtering and interactivity.

The bar graph displays state-specific drug-related deaths for a selected year. Position is again the primary visual channel, with the x-axis representing states and the y-axis indicating the number of drug-related deaths. Bars were chosen as the mark type because they effectively represent discrete values for categorical data. To maintain simplicity and focus on bar height (magnitude), a single consistent color was used across all bars. While this sacrifices the ability to highlight outliers through color, it reduces cognitive overload and keeps the graph clean. The x-axis uses state abbreviations instead of full names to accommodate all states neatly within the space. Users can explore temporal changes by interacting with a year slider, which updates the visualization dynamically. Tooltips provide exact death counts for each state, supplementing visual interpretation with precise numerical values. Additionally, the y-axis adjusts dynamically to the maximum value for the selected year, ensuring efficient use of visual space. Trade-offs include the choice of alphabetical sorting for states, which prioritizes organization over emphasizing trends, though alternate sorting methods (e.g., by death count) could provide additional insights.

Across both visualizations, no transformations like log scales were applied, as the raw data values effectively conveyed the desired scale and granularity. Every design decision was guided by the principle that "every pixel must be justified," ensuring marks (lines, bars) and channels (position, color) were carefully selected to enhance clarity, reduce noise, and support user-driven exploration of the data. The interactive features in both visualizations ensure a balance between simplicity and depth, allowing users to gain insights without being overwhelmed.

Part C: Interaction Overview

The interactive elements in the visualizations were designed to enhance usability, empower exploration, and provide insights tailored to user needs. The guiding design principle for interactions was to maintain simplicity while offering depth, ensuring that users could intuitively discover and use the features without requiring extensive instructions. The interactive features include filtering, tooltips, and dynamic updates for both visualizations, as well as a year slider for temporal exploration in the bar graph. These choices were informed by the nature of the dataset and the need to explore complex, multi-dimensional data in a clear and engaging way.

For the line graph, filters were implemented to allow users to focus on specific demographic groups or drug types, reducing visual clutter when analyzing detailed trends. By using checkboxes or dropdown menus, users can toggle lines on and off, providing control over the data displayed. These filters were chosen to address the common user need to compare subsets of data (e.g., between sexes or age groups). Dynamic scaling of the y-axis ensures that the graph remains readable and precise, regardless of the data subset chosen. Additionally, interactive tooltips appear on hover, providing precise data values for each point on the line. This feature helps users gain exact information without overcrowding the visualization. These interactions are immediately discoverable, as tooltips activate naturally during exploration, and filters are prominently displayed near the graph, making them easy to find.

In the bar graph, a year slider was introduced to allow users to explore how state-specific drug-related deaths change over time. This interaction was chosen because temporal trends are a key focus of the dataset, and a slider provides a simple and intuitive way to navigate through the years. Tooltips on each bar reveal the exact death count for that state in the selected year, providing additional precision to complement the visual representation. The bars dynamically update in response to the year slider, with smooth transitions ensuring users can easily follow changes in the data. The design decision to include a slider rather than static graphs for each year was driven by the need to save space while offering a seamless temporal exploration experience.

To make these interactions discoverable and usable, all interactive elements are prominently placed near or within the visualizations. Filters and the year slider are intuitively labeled, ensuring users understand their purpose without prior instruction. Visual feedback, such as the immediate update of graphs upon filter or slider interaction, reinforces the usability and responsiveness of the system. To make the experience interesting, the dynamic and responsive nature of the graphs ensures users can uncover patterns and insights through exploration, fostering a sense of engagement and discovery.

The process of designing these interactions involved identifying key user needs, such as the ability to explore subsets of data or temporal trends, and ensuring the interface remained uncluttered and approachable. Each interactive element was designed with the goal of balancing simplicity, functionality, and the overall aesthetic of the visualization, ensuring that the interactions enhance the user's ability to interpret the data without overwhelming or distracting them.

Part D: The Story

The primary goal of our project is to provide an interactive and intuitive visualization that enables users to explore drug-related deaths in the United States over the past few decades. Using dynamic line and bar graphs, users can analyze trends over time while filtering the data by demographics and drug types, allowing them to uncover how different populations and substances contribute to this ongoing crisis.

One of the surprising insights revealed by the visualization is the sharp increase in deaths linked to synthetic opioids in recent years, highlighting their significant role in escalating the epidemic. This trend underscores how certain drug categories have become dominant drivers of fatalities over time, emphasizing the need to address these substances specifically. Additionally, the visualization reveals disparities across demographic groups and states, with some regions and populations being disproportionately affected, offering an opportunity to identify areas that may benefit from targeted interventions.

The key insights we aim to convey through our visualizations include the importance of understanding trends in drug-related deaths over time and across different states to grasp the true scope and impact of the crisis. By leveraging filters for demographic groups and drug types, users can gain a deeper, more nuanced understanding of how specific populations and substances shape the epidemic. Our visualization also strives to emphasize the alarming rise of synthetic opioids and their outsized influence on drug-related mortality, urging policymakers and stakeholders to prioritize strategies for mitigating their impact. Through these insights, we hope to foster a more informed discussion about the crisis and inspire action to address its multifaceted challenges.

Contributions

Peipei: Organized Zoom meetings to facilitate discussions and actively address group tasks, while completing the majority of the report content collaboratively. Maintained a high level of coordination.

Lily: Developed project concepts and prompts, sourced relevant datasets, organized and facilitated team meetings, led discussions, and ensured effective communication to maintain project momentum. Designed and created the initial visualization, established the overall style and CSS for the project, authored summaries for both visualizations and provided constructive feedback and critique throughout the process. For me, figuring out the project topic and extracting valuable prompts from the raw dataset took the most time.

Haneul: Organized and planned team meetings and drafted milestones to keep the project on track. I actively participated in all team discussions, contributing to decision-making processes. I merged the final files together, fixing final errors, ensuring they aligned with the team's vision for functionality and aesthetic appeal. Additionally, I contributed to drafting the report, synthesizing team efforts and insights, and providing edits and feedback to improve clarity and cohesion.

Katie: Coordinated and scheduled team meetings while establishing milestones to ensure the project remained on track. Actively engaged in team discussions, ensuring effective communication. Designed the second visualization and helped merge the final files together, cleaned the second data set, and resolved errors.