**Project Name:** Micro-Migrations in NYC

**Rui Shen** [rshen65@garech.edu](mailto:rshen65@garech.edu) Conceptualization, data collection, data processing, map design, web development, sketches, and documentation

**Chunlan Wang** [cwang932@garech.edu](mailto:cwang932@garech.edu) Conceptualization, sketches

***Note: Due to slow data loading, it may take some time to wait for the page to load.***

**Introduction**

The initial motivation for wanting to present the "migration" behavior on the map may stem from our own constant "migration." We hoped to find a smaller entry point into this topic, beyond the larger perspective of global population migration, through either the data itself or our angle of data observation.

Initially, we considered documenting bird migrations but faced data shortages. Our focus then shifted to artworks, aiming to trace the movement of pieces across various collectors, museums, and institutions. However, only a few institutions' public datasets contained such information, with a broad time range, complicating effective presentation.

Ultimately, we chose taxi operation records due to the accessibility of the data and the ease of presentation, aiming to depict the small-scale daily migratory behavior within cities facilitated by taxis. When beginning to create the map, we posed the following questions:

1. Whether such migration exhibits certain temporal and spatial patterns.
2. Whether such migration is influenced by factors such as weather.

**Data**

Our original data comes from NYC OpenData, as few cities make their taxi operation records public.

The original data includes:

1. **2021 Yellow Taxi Trip Data:** <https://data.cityofnewyork.us/Transportation/2021-Yellow-Taxi-Trip-Data/m6nq-qud6/about_data> This dataset contains 30.9 million rows and 18 columns, with each row representing a taxi trip record. It details pick-up and drop-off locations, distance, fares, and more.
2. **NYC Taxi Zones:** <https://catalog.data.gov/dataset/nyc-taxi-zones>This dataset provides specific geographical information on New York taxi zones.
3. **Google Map API:** This data source aids in better presenting information for each taxi zone.

Data processing involved:

1. **Data Cleaning:** Removing obviously unreasonable data, such as non-existent taxi zone codes.
2. **Data Categorization:** Grouping the data by season, whether it's a weekday, and hourly intervals.
3. **Data Calculation:** Computing the population exchange data between different taxi zones.

**Maps**

**Interfaces：**

Our structure consists of two main interfaces.

**The first is dedicated to visualizing taxi operation data that has been categorized.** Users select time information and the type of data for visualization, with data types divided into two categories: pickups from the area and drop-offs to the area. This approach aims to comprehensively represent the directional nature of taxi data. The size of the data is indicated by opacity. Additionally, an information box is activated upon mouse hover to present basic details of different areas.

**The second interface visualizes micro-population migrations in New York City, facilitated by taxis.** Upon selecting a specific area, the map displays population exchanges with other areas, where red signifies inflow from the selected area to the current area and blue represents outflow from the current area to the selected area. The depth of the color indicates the volume of exchange. It is recommended to activate the timeline's play button after selecting an area to observe micro-migrations throughout a day. Clicking on another area changes the current area of focus. To return to the first interface, clicking on the selected area again will switch back, allowing for a smooth transition between interfaces.

**Components:**

1. **Season Selection:** Users can switch between taxi data for four seasons, with fill colors changing to symbolize different seasons as they switch.
2. **Date Type Toggle:** Users can toggle taxi data between weekdays and weekends, while the map's hue shifts between light mode and dark mode.
3. **Hover Information Box:** When users hover their mouse over different taxi zones on the interface, an information box is activated in the top right corner of the window. It displays the name, ID, average pickup and drop-off distances and costs, along with a landmark picture to give users a general impression of the area.
4. **Timeline:** Users can drag the timeline to switch between taxi data for different hours. The timeline is set to be playable; clicking play will move the timeline forward at a speed of one hour per second from the current position.
5. **Data Selection:** Users can select different data to be displayed on the map through a selection box in the interface.

**Limitation**

The current stage of this project has **numerous** shortcomings, largely due to time constraints and limitations in capability. Many envisioned effects have not been realized, and several structural issues exposed during the process remain unresolved for the time being.

1. **Data Authenticity**: Late in the map development phase, we realized that even after data cleansing, some erroneous data remained. These errors were mainly manifested in excessively long or short operation distances, which could be explained as mistakes in not turning off the meter or as extremely rare real events. However, lacking the capacity to judge, we had to include them in the final data presentation.
2. **Legend**: Due to difficulties in ensuring the authenticity of the data, especially regarding operation distances, we ultimately decided against including a legend in the first interface. Instead of interpolating for continuity in opacity representation, we opted for manually controlled median points.
3. **Comparison**: Our initial motivation was to identify patterns or influencing factors, making comparison a useful tool. We had planned to allow users to select two different times, which could be time intervals, seasons, or types of dates for comparison, but this was not implemented.
4. **Analysis**: Although data presentation should ideally avoid the creators' biases, we believe some analysis tools, including charts, should be introduced.
5. **Medium**: The current medium of presentation is neither very intuitive nor rich. Using different colors to delineate migration scenarios was just one strategy we initially envisioned. Our conceptual strategies also included connections marked by different colors and line widths, three-dimensional surfaces formed by point clouds, and the use of varying sounds to represent exchanges. However, these were not compared or integrated.