Final Project Report: Direction-Based Pattern Map & Flow Visualizations for NYC Taxi Trips

Students: Mehran Mohammadi – 51842181 / Mookantsa Kamogelo Dorcus – 12334162

Abstract

This report presents the reconstruction and extension of spatial interaction pattern visualizations for the New York City Yellow Taxi dataset using H3 hexagonal grids, direction-based aggregation, and flow map rendering. The system integrates three complementary views into a single-file web application: a paper-faithful direction-based pattern map (FIX4), an exploratory main pattern map, and a global overview flow map. The workflow includes multi-million-trip preprocessing, directional binning, normalization, and the creation of reusable JSON data products to support interactive analysis and presentation.

Contents

1 Introduction

This project was conducted as part of the course **193.166 Visualisierung (VU 4,0), Winter Semester 2025**. The work reproduces and extends the methodology proposed in the paper *Visualizing Spatial Interaction Characteristics with Direction-Based Pattern Maps* by Yao et al., applying it to the New York City Yellow Taxi Trips dataset.

This report presents the reconstruction, extension, and visualization of spatial interaction patterns in the New York City Taxi dataset using H3 hexagonal grids, direction-based pattern analysis, and flow map visualization.

The system includes an additional **paper-faithful Direction Pattern tab (FIX4)** that closely reproduces the glyph design, aggregation logic, and visual encoding described in the reference paper, while also providing interactive extensions such as click-to-inspect incoming and outgoing flows per hexagon.

2 Dataset Description

2.1 Source

The dataset consists of NYC Yellow Taxi trip records. After preprocessing, approximately **4.2 million valid trips** were retained. Each trip contains pickup and drop-off LocationIDs, geographic coordinates (via Taxi Zone centroids), trip distance, and timestamps.

2.2 Sampling

For overview-style flow visualization, a **1% random sample** of trips was extracted (roughly 42,000 trips). This sample is used exclusively for the overview flow map to provide city-scale context at low rendering cost.

3 Preprocessing & Data Engineering

3.1 Coordinate Mapping

Taxi Zone LocationIDs were mapped to latitude/longitude coordinates using centroid lookup tables. Trips with invalid or missing coordinates were removed.

3.2 H3 Hexagon Assignment

Pickup and drop-off coordinates were assigned to **H3 hexagons at resolution 8** (approximately 1 km cell size), aligning with the spatial scale used in the reference work.

3.3 Direction Computation

For each trip within a hexagon, a bearing was computed and classified into **six directional bins** (e.g., N, NE, E, SE, SW, W). For each hexagon and direction, the following were aggregated:

- Trip count
- Total distance
- Mean distance

3.4 Normalization

Two normalized metrics were derived:

- **count_norm**: relative magnitude of trips per direction within a hexagon
- **distance_norm**: normalized mean distance per direction

3.5 Final Data Products

The preprocessing pipeline produces three complementary data products, each serving a distinct role in the visualization framework:

- **hex_directions_real.json** — Per-hexagon directional aggregates derived from the full dataset. For each active H3 cell (e.g., 138 hexagons), trips are aggregated into six directional bins storing metrics such as trip count and mean distance. This file is the analytical backbone of the direction-based pattern maps.
- **trips_sample.json** — A 1% random sample of origin–destination taxi trips *without* hex-level enrichment. It is used for the global overview flow map to provide a lightweight city-scale representation of movement

patterns.

- **trips_sample_with_hex.json** — A sampled OD trip file that additionally includes pickup and drop-off H3 hex IDs. This enables direct linkage between OD flows and hexagonal spatial units, supporting hex-aware flow inspection and interactive per-hex highlighting in the direction pattern views.

4 Visualization Framework (Three Complementary Tabs)

All visual outputs are integrated into a **single-file, three-tab web application**, supporting reproducibility and ease of demonstration.

4.1 Direction Patterns (Paper-Faithful — FIX4)

Purpose: Visualize directional magnitude and directional distance inside each hexagon using a dual-layer hexagonal glyph.

Visual encoding:

- Outer mesh: H3 cell boundary
- Inner layers:
  - Gray scale → magnitude (trip count or normalized count)
  - OrRd scale → distance (mean distance)
- Each hexagon is subdivided into **six triangular sectors**, one per direction
- Color classes computed using **Jenks natural breaks**
- Distance values can be capped at **p95 / p98 / p100** to reduce outlier influence

Interaction:

- Minimum-trips slider filters low-activity hexagons
- Direction dropdown filters directional sectors
- Radio controls switch magnitude metric and distance cap
- Clicking a hexagon reveals **incoming/outgoing flow lines** for that hexagon

4.2 Hex Intensity (Main Pattern Map)

Purpose: Provide an exploratory and presentation-oriented view of directional flows with strong emphasis on readability.

Features:

- Hexagons colored by total trip volume
- Directional arrows indicating flows by direction and magnitude
- Optional animated dashed arrows
- Story modes: hex-only view, all directions static, medium traffic, high traffic
- Export: PNG and SVG

4.3 Global Flow (Overview Map)

Purpose: Provide geographic context and a macro-scale view of OD patterns.

Features:

- NYC borough boundaries (GeoJSON)
- H3 grid overlay
- 1% trip sample rendered as low-opacity OD lines
- Zoom/pan interaction
- Export: PNG

5 Comparison with the Reference Article

5.1 Matched Components

- Hexagonal spatial tessellation
- Six-direction classification
- Aggregated directional statistics per spatial unit

- Distance and magnitude encoded via color and geometry
- Flow overview map for contextual understanding

5.2 Enhancements Beyond the Paper

- Interactive filtering (direction, thresholds, caps)
- Click-to-inspect flows per hexagon
- Story modes for structured presentations
- Downloadable publication-quality images
- Single-file deployment for reproducibility

5.3 Differences

- H3 indexing is used to ensure reproducible hexagonal tiling
- Mean distance is used as a distance summary per direction
- Implementation emphasizes lightweight web deployment and interactive analysis

5.4 Paper-Faithful Direction Pattern Tab (FIX4)

The FIX4 tab explicitly reproduces the glyph design and visual encoding described in the reference work. It separates magnitude and distance into two concentric hexagonal layers, applies Jenks natural breaks for classing, and introduces selectable distance caps (p95, p98, p100). Clicking a hexagon reveals its incoming and outgoing flows, extending the paper's static design into an interactive analytical tool.

6 Methodology (Step-by-Step Pipeline)

6.1 Data Acquisition

NYC Yellow Taxi Trip Records were obtained from a public source. NYC boundary GeoJSON files and Taxi Zone centroid lookup tables were collected to support geographic mapping and contextual overlays.

6.2 Data Cleaning & Preprocessing

1. Load trip data (e.g., Parquet format).
2. Remove rows with missing or invalid pickup/dropoff LocationIDs.
3. Map LocationIDs to geographic centroids.
4. Filter invalid coordinates and out-of-bound locations.
5. Compute trip distances (e.g., Haversine great-circle distance).
6. Compute initial bearing between pickup and drop-off.
7. Classify bearings into six direction bins.
8. Assign pickup and drop-off coordinates to H3 hexagons (resolution 8).
9. Aggregate per-hex totals and per-direction statistics.
10. Normalize magnitude and distance metrics.
11. Export JSON data products for visualization.

6.3 Data Structures Used

- Hex-level directional JSON driving the analytical visualizations.
- 1% OD trip samples for overview flow rendering.
- Hex-enriched OD samples enabling interactive flow inspection and linking.

7 Evaluation & Discussion

7.1 Performance

Pre-aggregation enables smooth interaction in the browser despite multi-million-trip input data.

7.2 Usability

Clear controls, tooltips, and tab separation support both expert analysis and non-expert exploration.

7.3 Visual Clarity

Opacity control, Jenks classification, and distance capping reduce clutter and outlier dominance.

7.4 Scalability

The framework can be adapted to other cities, different spatial resolutions, and larger datasets.

8 Limitations

- Taxi Zone centroids approximate origin and destination locations, not full GPS trajectories.
- Directional aggregation abstracts individual route geometry.
- High-density areas may still exhibit partial occlusion in flow overlays.
- Temporal slicing and time-dependent pattern maps are not implemented in the current version.

9 Future Work

- Temporal pattern maps (hourly/daily) and time-based filtering
- WebGPU rendering for higher performance at scale
- Directional clustering and anomaly detection
- Integration of land-use or demographic layers for richer interpretation

10 Conclusion

The updated system integrates three complementary visualizations—Direction Patterns (FIX4), Hex Intensity, and Global Flow—into a single coherent analytical framework. By combining H3 spatial indexing, direction-based aggregation, and interactive D3.js visualizations, the project reproduces the reference paper while extending it into a flexible, presentation-ready analytical tool.