



# Taxi Driver Behavior Pattern Mining Based On Exploratory Data Analysis



Qian Xie

Supervisor: Yang Yu (Assistant Professor)

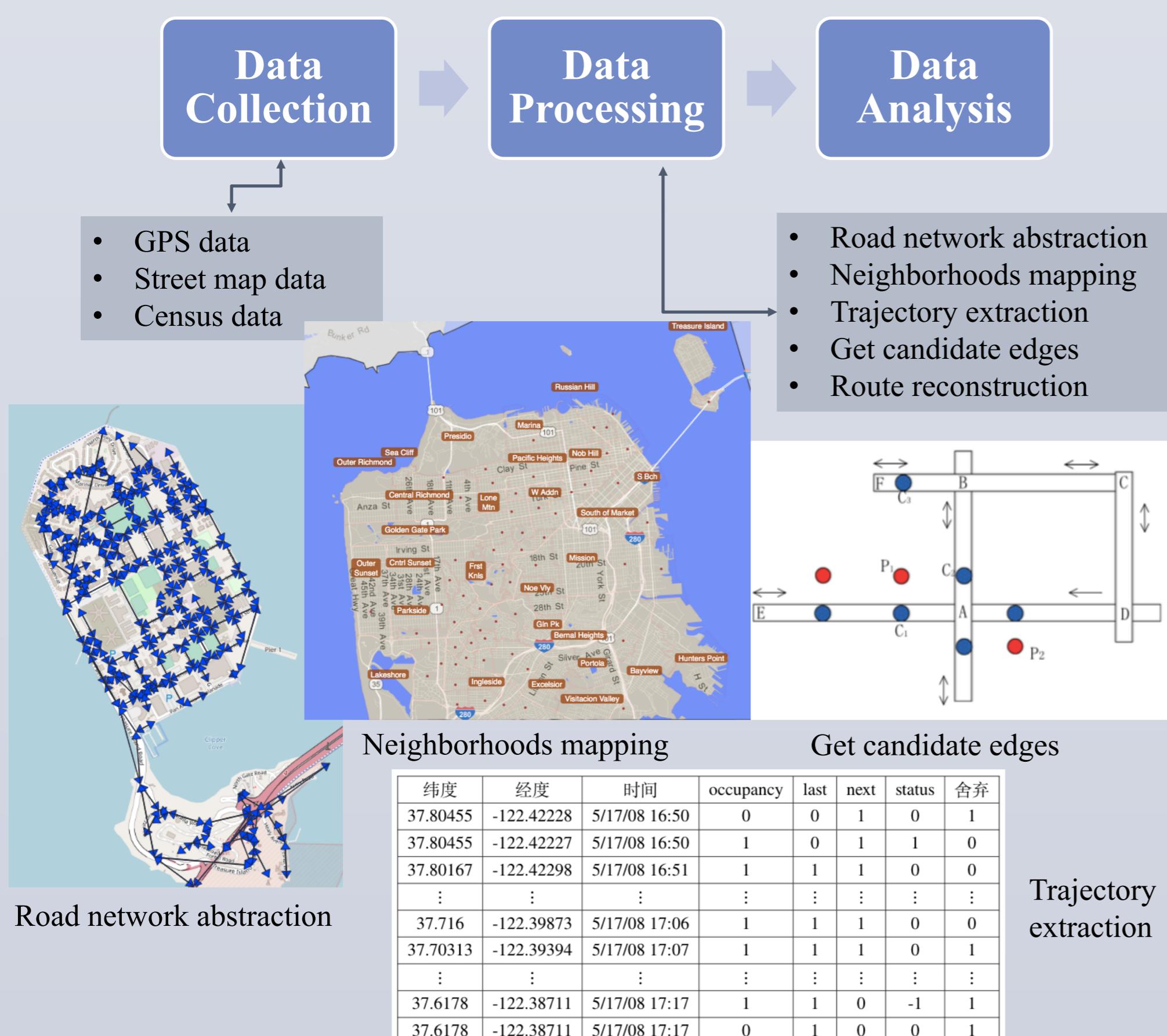
## ❖ Abstract

- Detour: deviation from the optimal path
  - Multiple candidate routes to choose from given OD
  - Optimal path can be the shortest one among them
  - Deviation is quantified by length difference and similarity measure
- Extract and reconstruct paths from taxi GPS raw data
  - Most current work only extract trajectories
    - Problem: GPS error and lack of information
  - Design multiple efficient algorithms to select the best-fit path
  - Apply various data structures and pruning techniques to speed up
- Apply EDA to study taxi driver detour behaviors
  - Relate GPS data with census data to provide new perspectives
  - Micro-level and macro-level analysis based on data hierarchy
    - A taxi has multiple trip records (paths)
      - Deviation distribution and correlation with demographic factors
    - The dataset consists of multiple taxi records
      - Use power law to fit the distribution and compare the difference

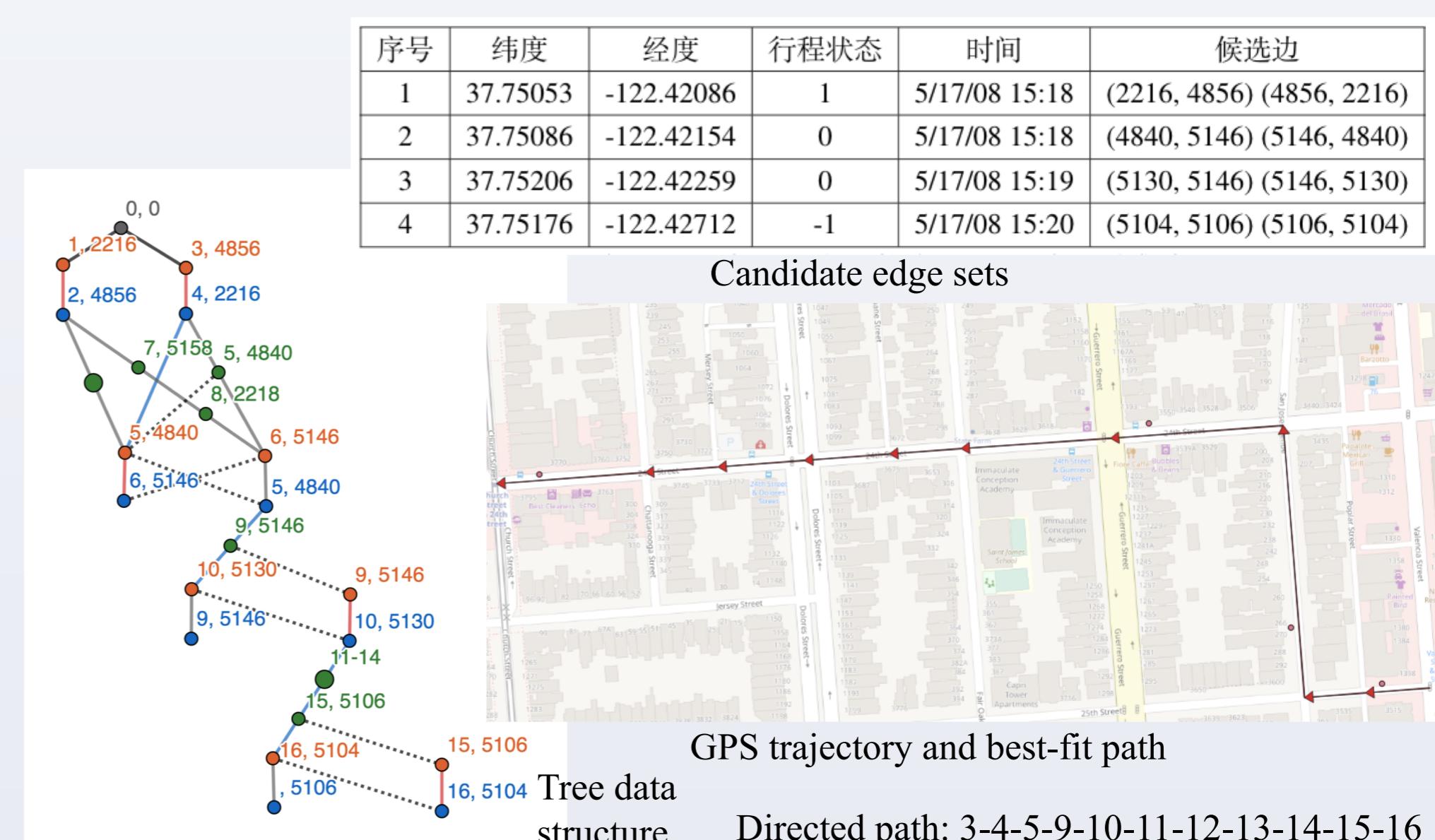
## ❖ Research Problem

- Best-fit path: a directed path that best describes how a taxi travels from origin to destination given a trajectory (series of GPS points collected from the taxi)
  - Given a directed graph  $G=(V, E)$  and a GPS trajectory  $P = [p_1, p_2, \dots, p_n]$ , where  $E$  consists of directed edges representing roads on the map
  - Goal: output directed path  $[e_1, e_2, \dots, e_k], e_i \in E$  satisfying
    - There exists matching edges  $[e_{j_1}, e_{j_2}, \dots, e_{j_n}]$  for  $[p_1, p_2, \dots, p_n]$  where  $1 = j_1 \leq j_2 \leq \dots \leq j_n = k$  where  $d(p_i, e_{j_i}) \leq \delta$ ,  $\delta$  is a pre-defined threshold
    - Denote  $p_i$ 's projection on  $e_{j_i}$  as matching point  $p'_i$ , then the sub-path from  $p'_i$  to  $p_{i+1}$  is the shortest
  - Assumptions: most GPS error not exceed  $\delta$ , driver locally chooses shortest path
- Detour behaviors: can attribute to poor sense of direction, avoidance of congestion, safety concern, despising the poor, race discrimination, etc.
  - Does the driver reveal the tendency to detour?
  - Whether the detours are related to certain demographic characteristics?
  - What are the similarities and differences among drivers' detour patterns?

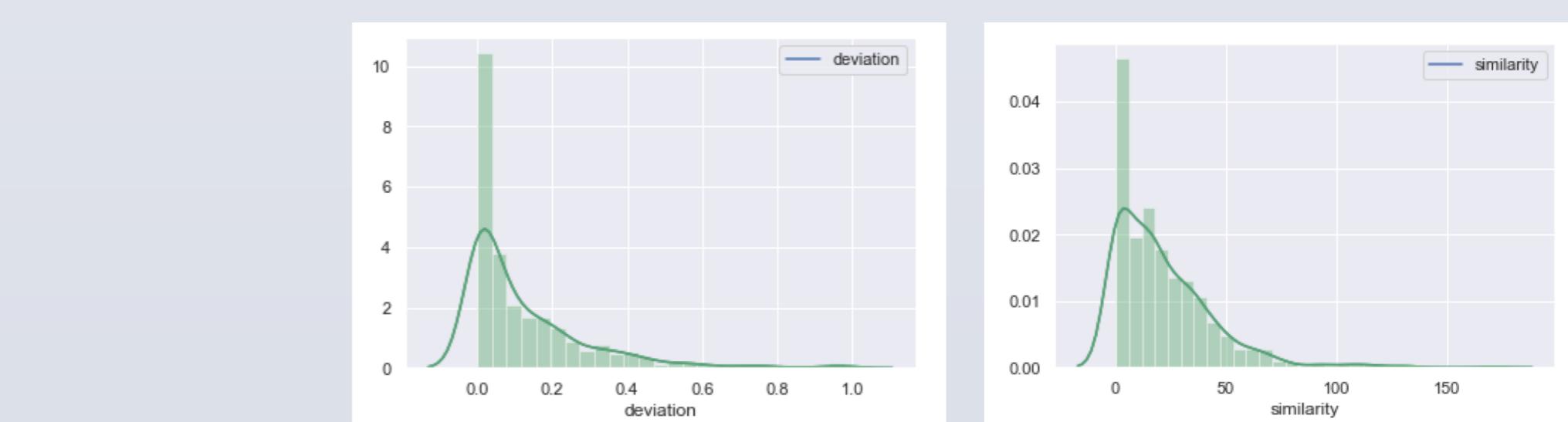
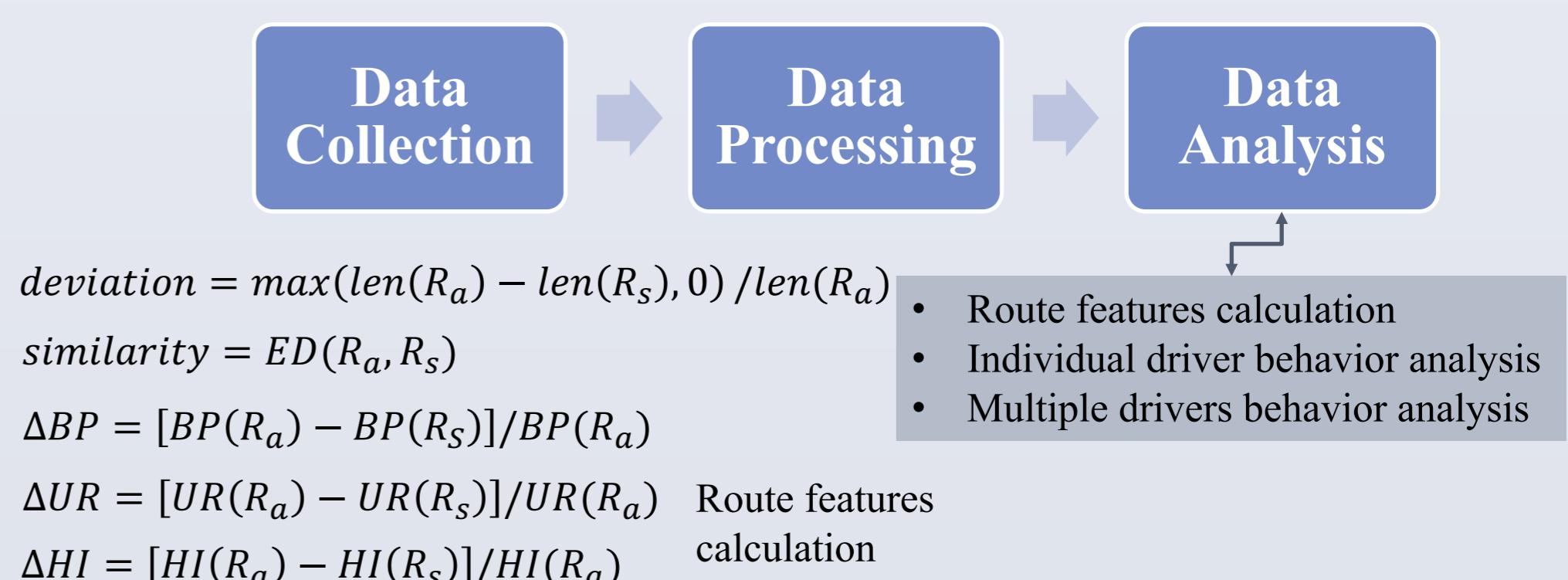
## ❖ Data Processing



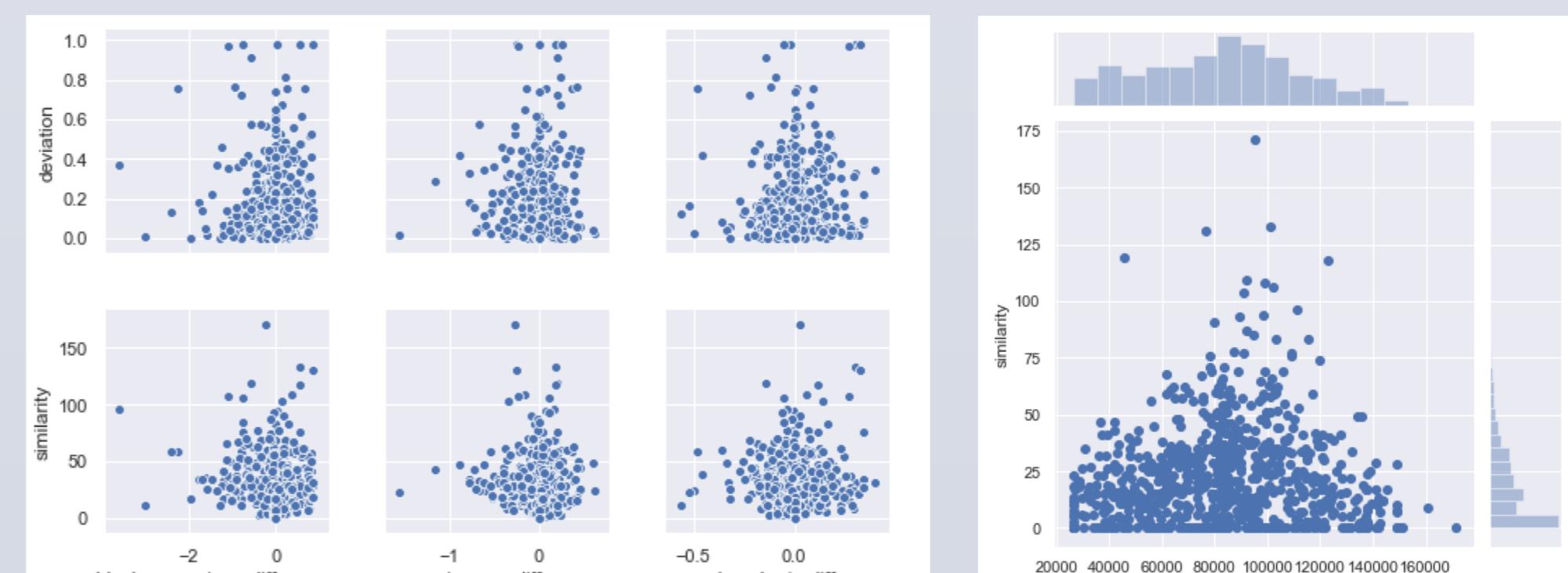
## ❖ Route Reconstruction Example



## ❖ Behavior Analysis



Both deviation and similarity are under long-tailed distribution



Correlation between detour behaviors and demographic characteristics

Detoured routes present more neutral attitude to demographic factors

## ❖ Future Work

- Observe if other drivers show similar correlation pattern and do further analysis
- Add speed and time period as new features
- Try to explore on causal relationship between detour behaviors and features including demographic characteristics