UTC Project Information	
Project Title	Privacy Risk Evaluation of Human Mobility Data for Urban Transportation Planning
University	University of Washington
Principal Investigator	Jan Whittington
PI Contact Information	janwhit@uw.edu
Funding Source(s) and Amounts Provided (by each agency or organization)	University of Washington PacTrans $15,000 University of Washington $15,000
Total Project Cost	$ 30,000
Agency ID or Contract Number	69A3551747110
Start and End Dates	September 16, 2020-September 15, 2022
Brief Description of Research Project	This project aims at developing privacy risk evaluation metrics for human mobility data that preserve both users' right to privacy and the utility of data for transportation planning and operations. Practitioners in the public and private sector have yet to resolve the pressing issue of how to condition (e.g., aggregate, suppress, modify) data to be shared or published to ensure its usefulness in analysis while preventing the re-identification of persons represented in the data. Built on literature of the re-identifiability of persons from mobility data (e.g., De Montjoye et al. 2013), the purpose of this project is to contextualize privacy risk in built environment and to place the resulting probabilities into a flexible and easy-to-use tool for public agencies. The project will estimate the tradeoffs between privacy protection and information loss to help public agencies select privacy preserving strategies for conditioning the data that will be optimized for neighborhood types (i.e., land use, density, property value) and user groups (i.e., age, gender, amount of total travel).

Describe Implementation of Research Outcomes (or why not implemented) Place Any Photos Here	The research outcomes have not been implemented yet. The current project is considered a pilot project using a regional level travel survey. To generalize the outcomes for implementations, the next step is to verify the outcomes from datasets based on other regions and include other common types of data, such as mobile phone tower data, used in the daily practice of public transportation agencies.
Impacts/Benefits of Implementation (actual, or anticipated)	The project has two anticipated implementations for data collection, processing, and publication practices. First, generalization or aggregation is a common strategy by public agencies to de-identify individuals in a dataset. However, it is often tricky to determine the optimal level and the right fields for aggregation as over-aggregation may lead to excessive information loss while under-aggregation may be insufficient to achieve a desired level of k-anonymity. By showing the structural variations in re-identification risk, this project suggests the possibility of de-identifying data according to variation in urban areas and population segments, which could help to reduce information loss and offer greater potential protection of the privacy of individuals whose data records are more unique than others. Secondly, the methodologies applied in this project can help detect and mitigate algorithmic biases in current data practices. This project found that travelers whose spatiotemporal traces are unique from others consisted of higher percentages of non-white persons and lower income populations, which may reflect the symptom of existing geographical inequality. Since these population groups also tend to be underrepresented in the data set, it makes them more vulnerable to the re-identification risk than others. When designing oversampling strategies to account for these underrepresented population groups, it is important to consider both the distribution of residential as well as travel location.
Web Links • Reports • Project Website	