



# Using Big Data to Measure Small Populations

Andrew Campbell, UC Berkeley  
[andrew.campbell@berkeley.edu](mailto:andrew.campbell@berkeley.edu)

*UC Berkeley Team:*

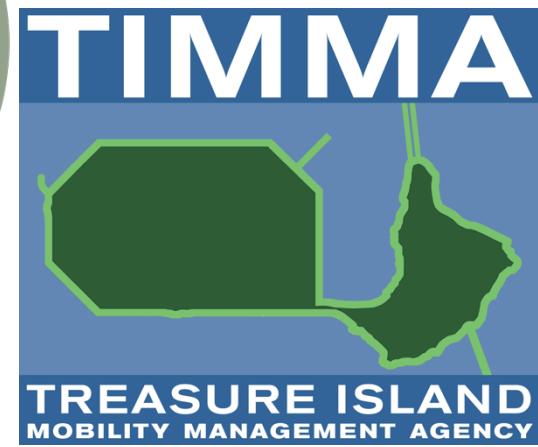
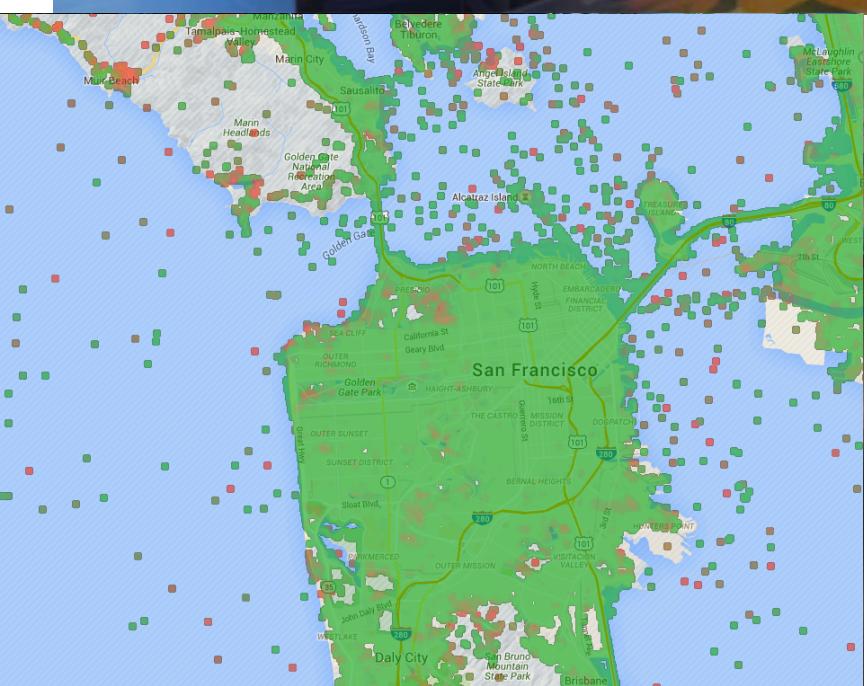
- Prof Alexei Pozdnukhov
- Mogeng Yin
- Side Fegin
- Madeleine Sheehan
- Danqing ZHANG

*In Collaboration With:*

- AT&T Labs Research
- SFCTA



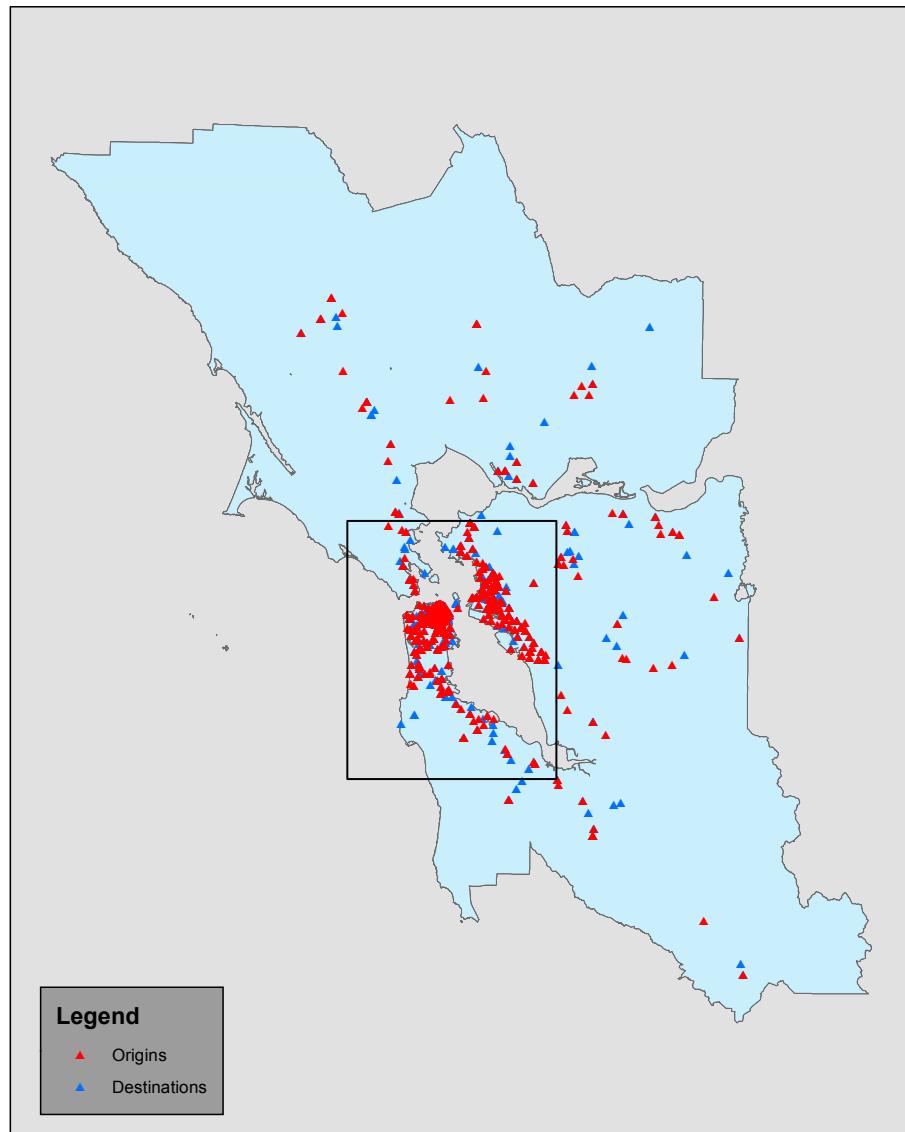
# Travel Data: Problems and Opportunities



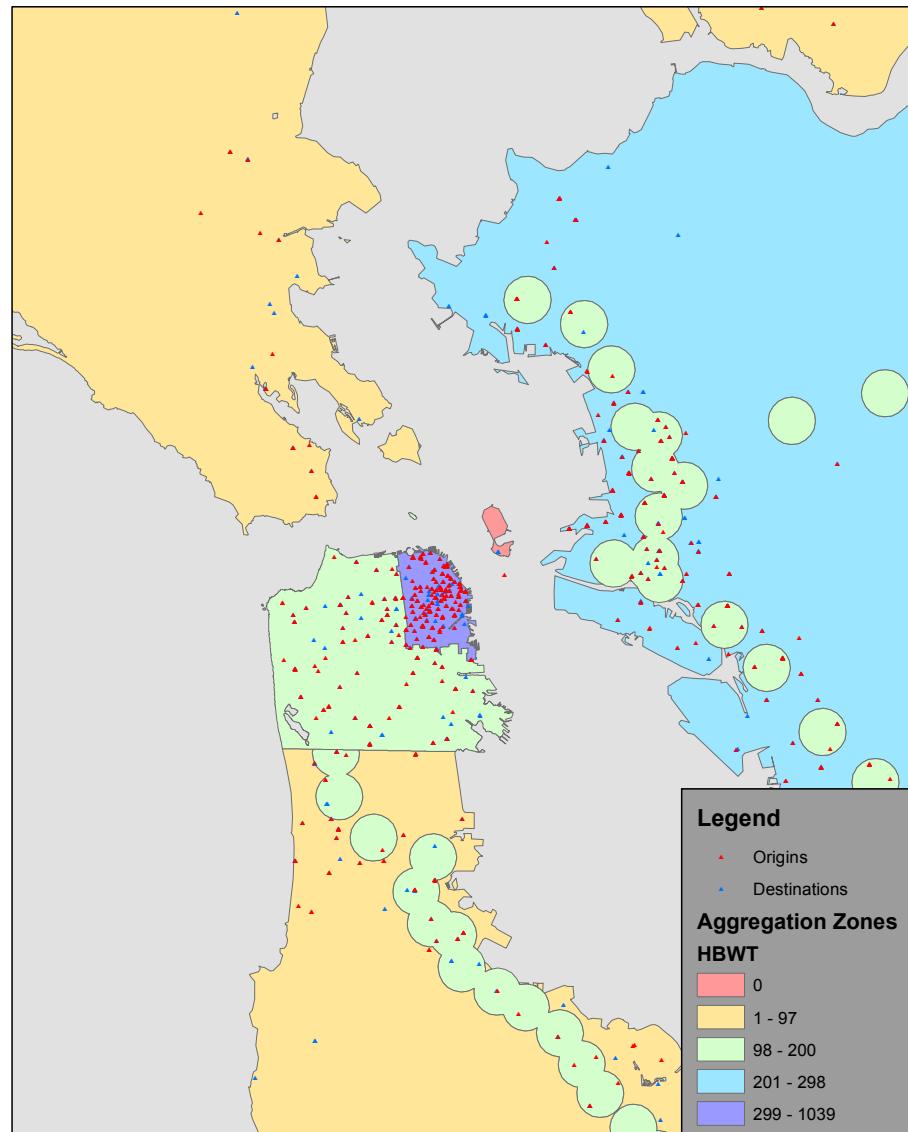


# Treasure Island - Spatial Scope and Aggregation

Spatial Scope of Mobility Study



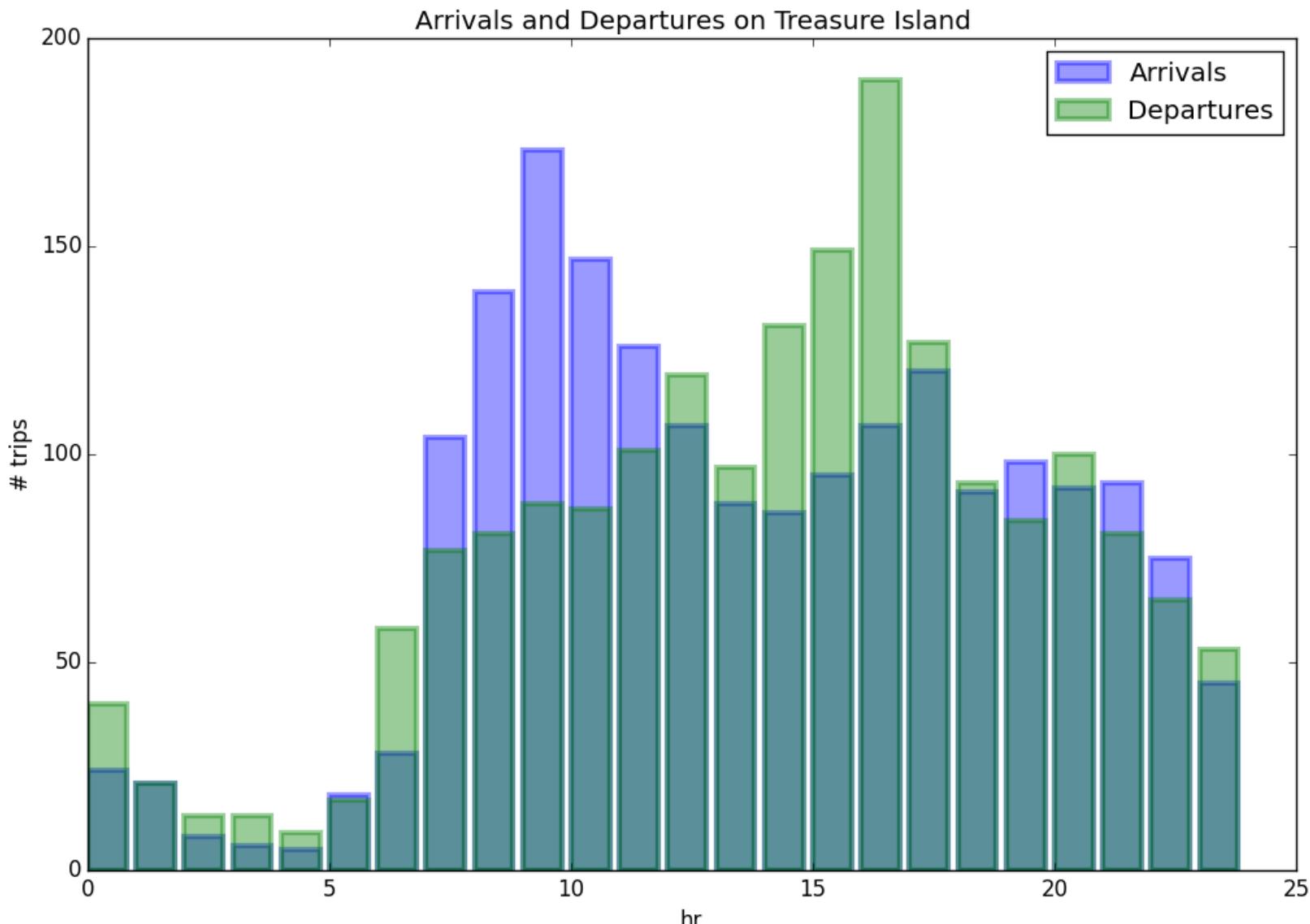
Aggregation Zone HBWT Counts



"No personally identifiable information (PII) was gathered or used in conducting this study. The mobility data that was analyzed was anonymous and aggregated in strict compliance with the carrier's privacy policy."

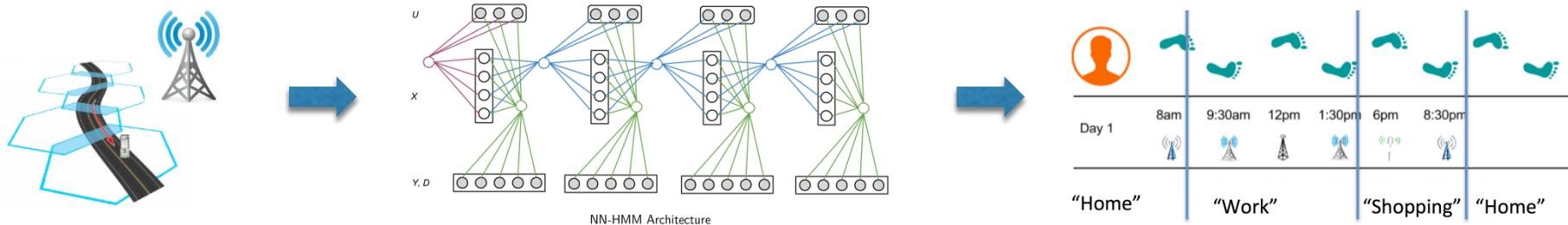


# Treasure Island – Temporal Distributions



*"No personally identifiable information (PII) was gathered or used in conducting this study. The mobility data that was analyzed was anonymous and aggregated in strict compliance with the carrier's privacy policy."*

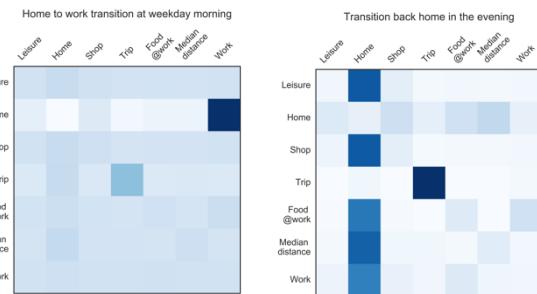
# Activity-based demand with machine learning



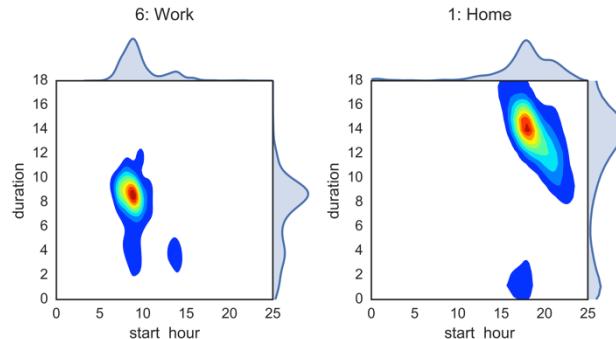
Human activity recognition from sequential location data with generative machine learning models:

- Recognize activity states (fit model structure with hidden variables)
- Learn daily activity pattern (conditional transition probabilities)
- Learn models for observable variables (destination choices and durations of activities given space/time context)

Transition probabilities



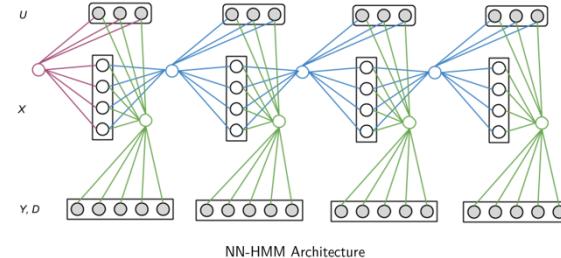
Activities durations



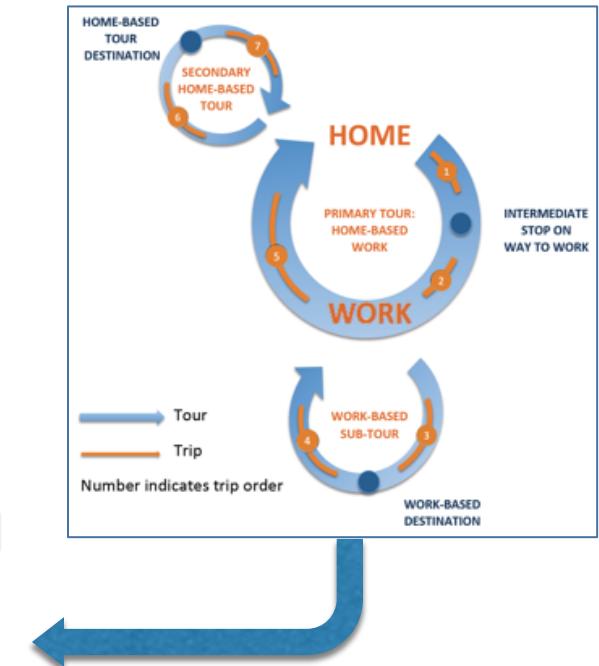
# The SmartBay – prototype for the Public-Private-Academic Partnership



San Francisco Bay Area mobility simulation from cellular data



Learn travel demand model for a population



Run regional mobility simulation scenarios

1M agents, 500 shown in visualization

