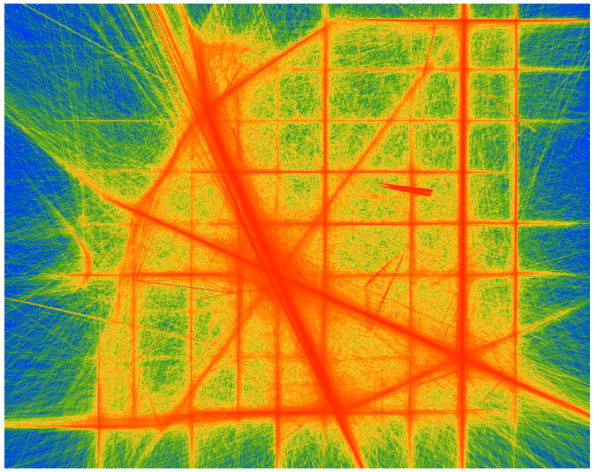
CENTER FOR GEOSPATIAL INFORMATION SCIENCE

# New opportunities through big mobility data analytics



KATHLEEN STEWART

CENTER FOR GEOSPATIAL INFORMATION SCIENCE

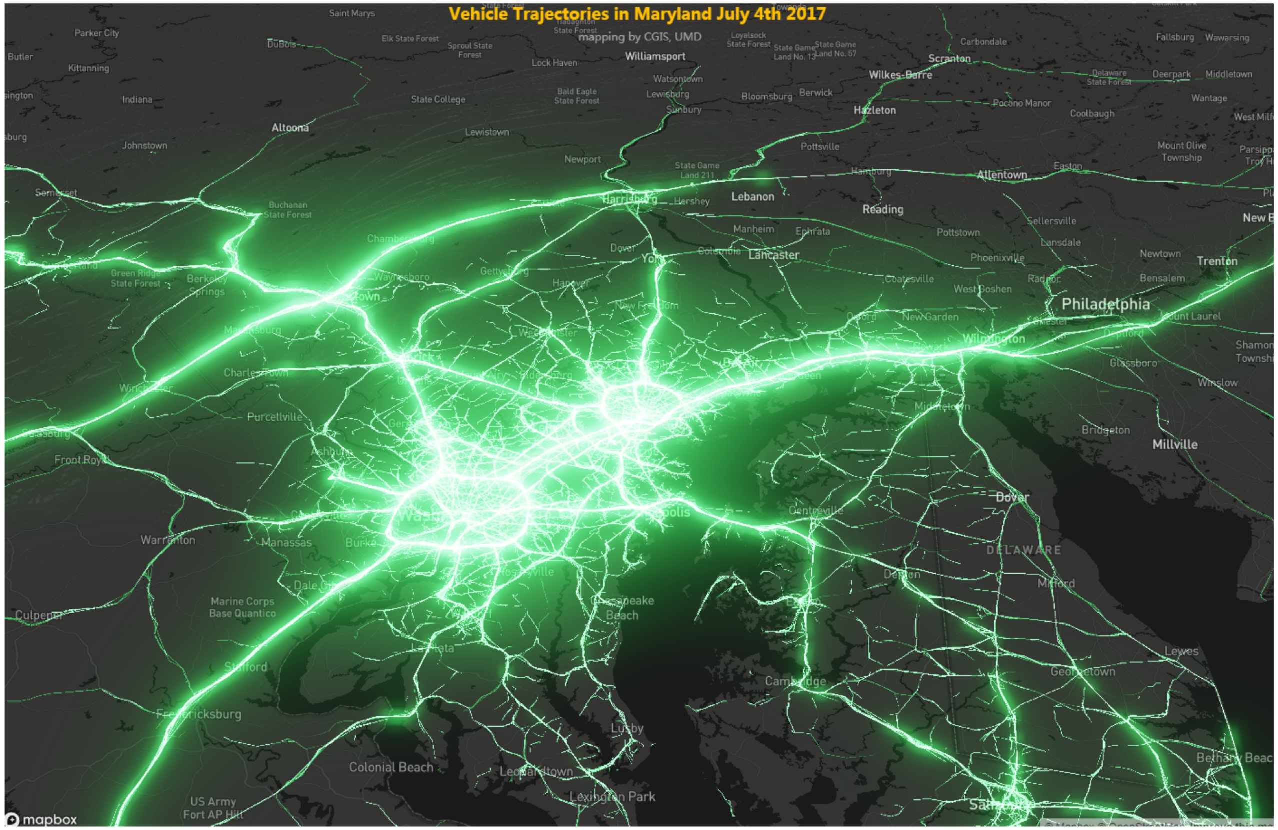
UNIVERSITY OF MARYLAND, COLLEGE PARK, MD

STEWARTK@UMD.EDU

*Location Powers: Data Science Summit, November 13, 2019*

# The opportunities: Vehicle travel patterns

50,282 trips for the 4th of July, 2015

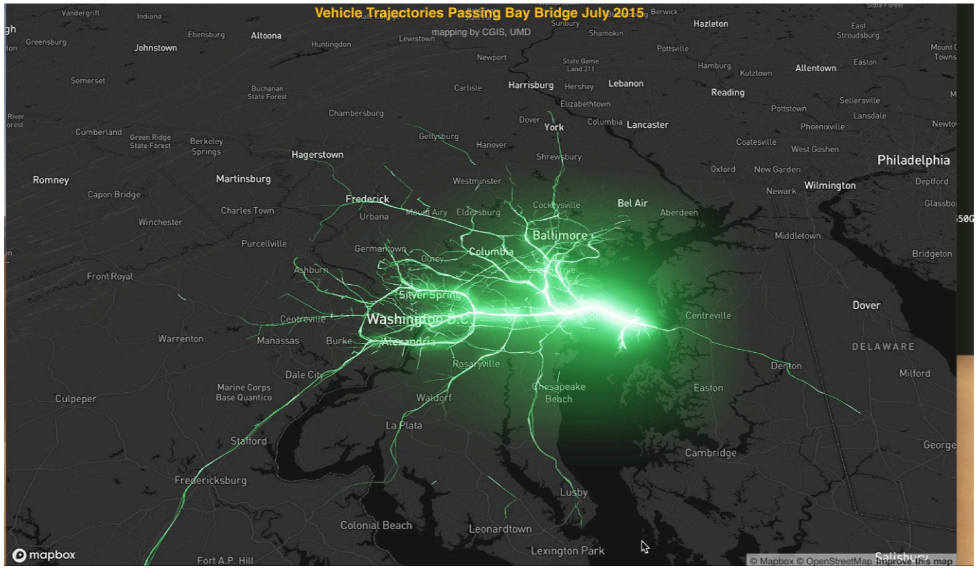


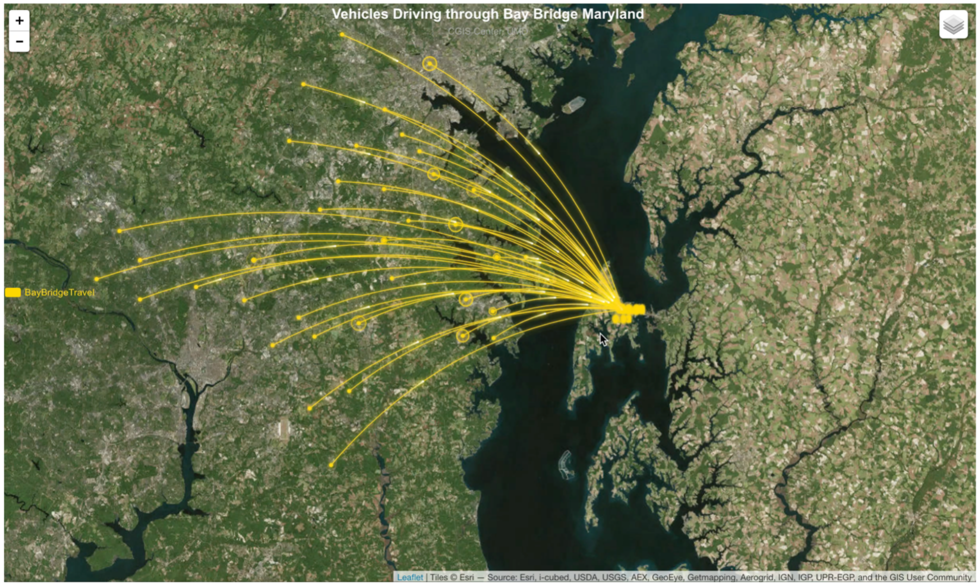

**J. Fan**

Especially useful for **critical or significant locations** such as a **key bridge** at a **key time (July 4th...)**

**Exposes different travel patterns over space and time**

**These 2 visualization tell different stories...**



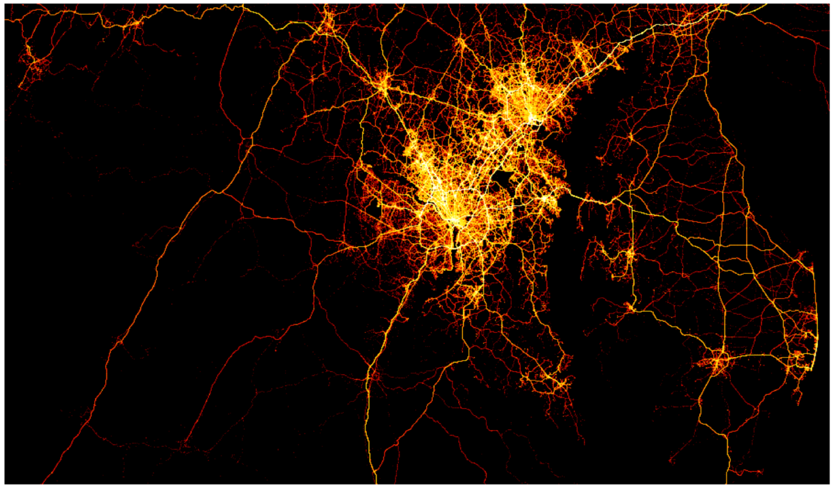
# Space-time patterns are important…

Data that give rise to such patterns are available from different data sources… **GPS waypoint** data, **cell phone** data, **location-based app** data, as well as **other sensors (fitness trackers)**…..

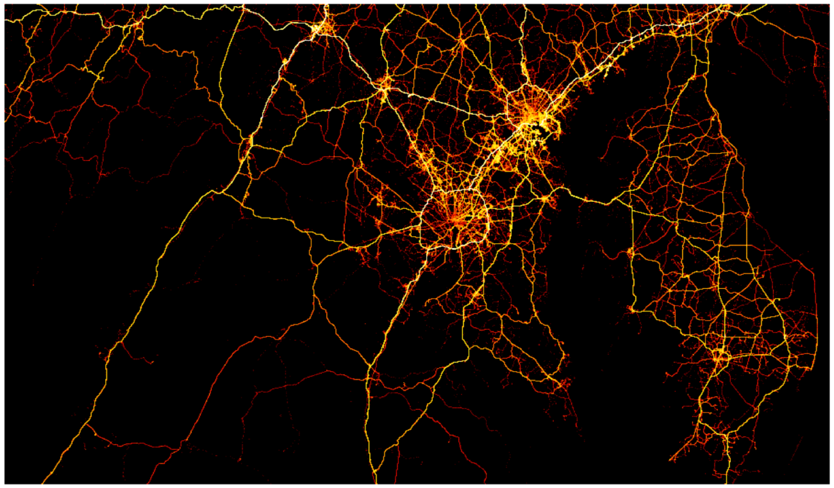
Here we see **big trajectory data** (GPS waypoints transformed into trajectories) useful for highlighting travel behaviors of different groups

We want to **expose different dynamic behaviors** over **space and time**

Important for:
- understanding urban/rural differences
- risk exposure
- Evacuation
- Multi-scale



Passenger vehicle trajectories, Maryland, 2015



Truck fleet trajectories

GPS data from in-vehicle sensors, Maryland, 2015
100 million waypoints, 5 million trajectories

## However, big GPS trajectory data comes with its own challenges…

- **Hundreds of millions** of **GPS trajectory data** have to be **matched** back to the underlying road network in order to determine the traffic volume pattern on different segments of the road network
- **Irregular GPS sampling intervals** and **gaps** exist for vehicles travelling on the road network derived perhaps from the fixed sampling intervals that can vary between devices

```
b4ba00bd9748981fc74c19afe5bacfc0,0,2015-06-01T10:53:00.000Z,40.1307,-77.0202,,
b4ba00bd9748981fc74c19afe5bacfc0,1,2015-06-01T10:56:00.000Z,40.1172,-77.0374,,
b4ba00bd9748981fc74c19afe5bacfc0,2,2015-06-01T11:01:00.000Z,40.0756,-77.0535,,
b4ba00bd9748981fc74c19afe5bacfc0,3,2015-06-01T11:01:12.000Z,40.0731,-77.0555,,
b4ba00bd9748981fc74c19afe5bacfc0,4,2015-06-01T11:06:00.000Z,40.0111,-77.1013,,
b4ba00bd9748981fc74c19afe5bacfc0,5,2015-06-01T11:11:00.000Z,39.939,-77.143,,
b4ba00bd9748981fc74c19afe5bacfc0,6,2015-06-01T11:16:00.000Z,39.8687,-77.1945,,
b4ba00bd9748981fc74c19afe5bacfc0,7,2015-06-01T11:21:00.000Z,39.7931,-77.2069,,
b4ba00bd9748981fc74c19afe5bacfc0,8,2015-06-01T11:26:00.000Z,39.7467,-77.2745,,
b4ba00bd9748981fc74c19afe5bacfc0,9,2015-06-01T11:29:04.000Z,39.7181,-77.3079,,
```

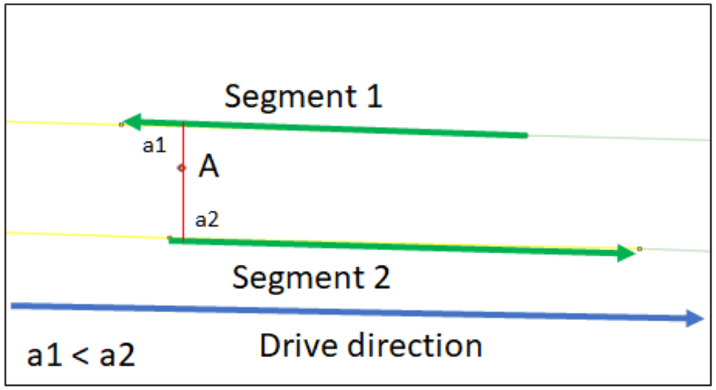
# Requires trajectory reconstruction algorithms

**Snap way-points of a trip to road segments**

- Find the K-nearest neighbors (k=10, and distance within 100 meters) as the candidates
- Select the candidate with largest cosine

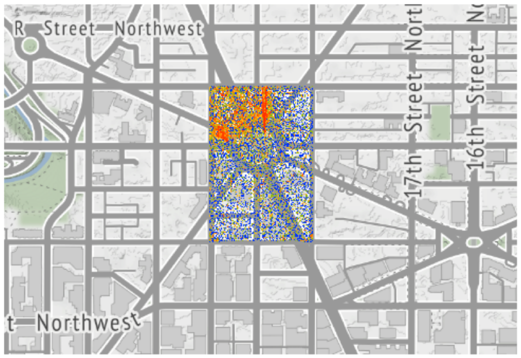
**Fill segment gaps by heuristic algorithms**

- Due to large time interval between two waypoints (1~2mins or longer), intermediate segments may be missing
- Use road network topology
- **Need to use a [shortest path] algorithm** to fill the gaps between two segments
    - This assumes that drivers take the shortest path between the two recorded waypoints…

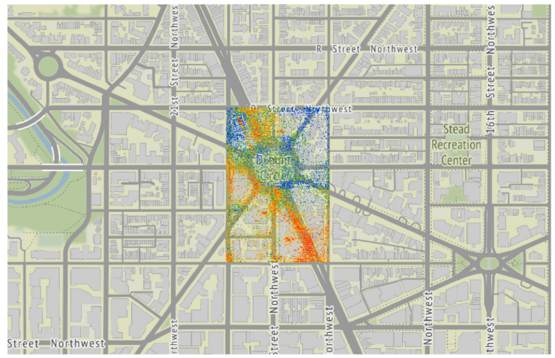



*Due to the uncertainty of GPS positioning, we cannot always employ the nearest segment to snap the waypoint*

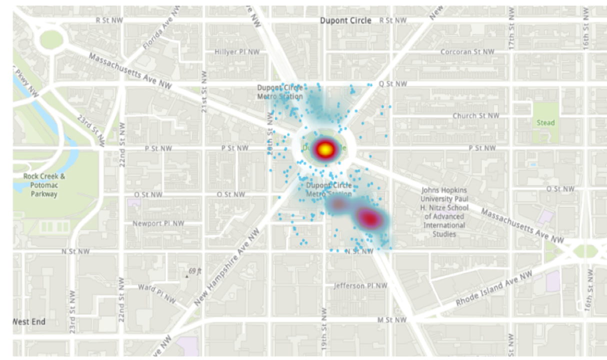
# Mobility data now comes from numerous different data sources….



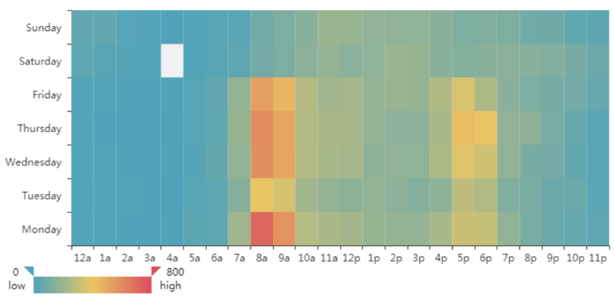

Cell phone data – ~23,775 sample points, ~14,000 devices

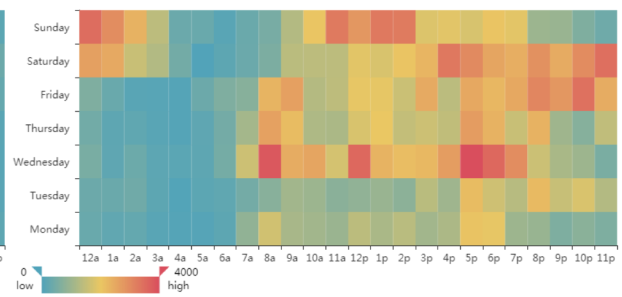


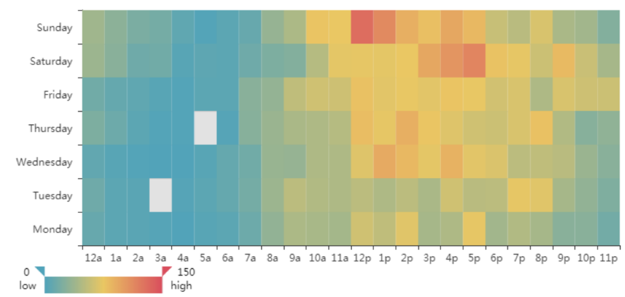

LBS data – ~240,000 sample points from ~7,300 devices




Geo-tagged tweets – ~7,000 sample points from ~2,600 individuals, 2017

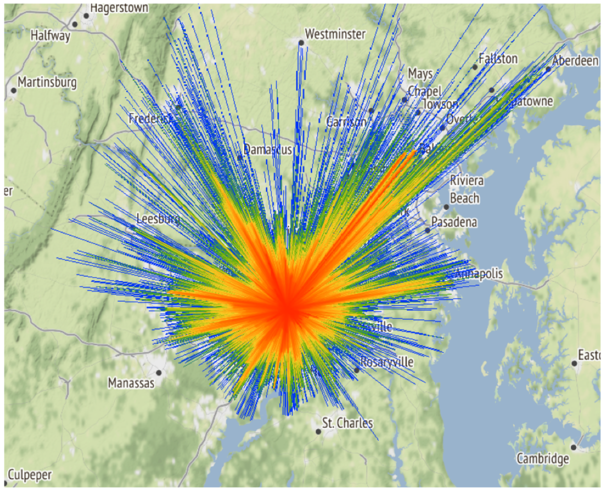


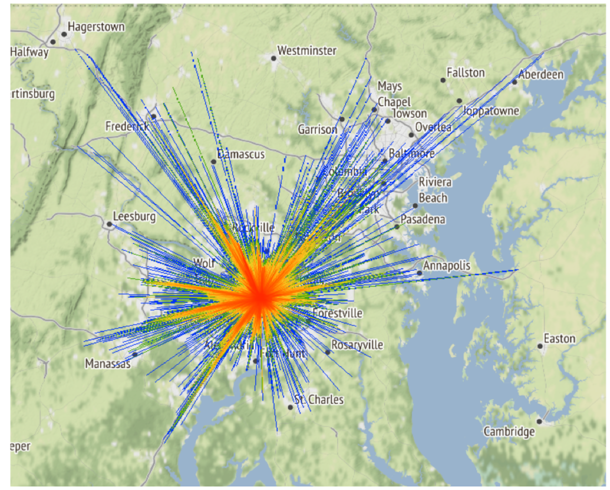

Dupont Circle, Washington, DC

# Trips originating from Dupont Circle (origin-destination analysis)



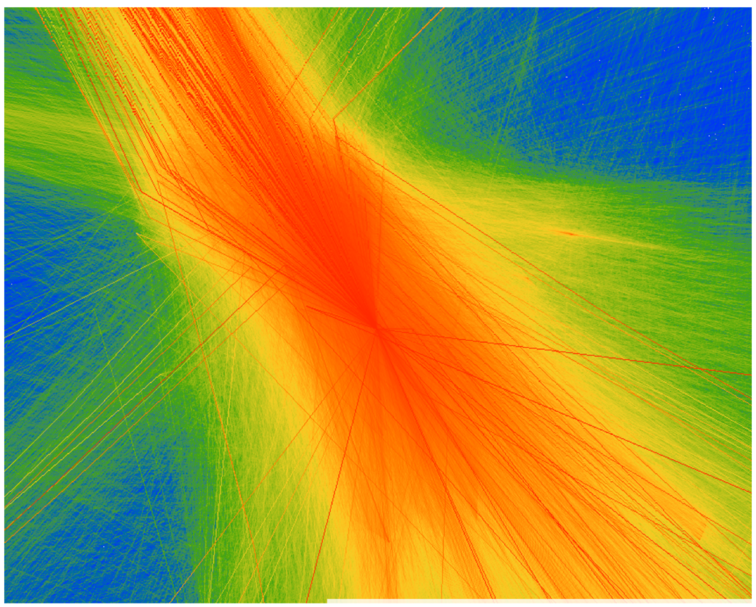

Cell phone data trips

More than 2,400,000 individuals



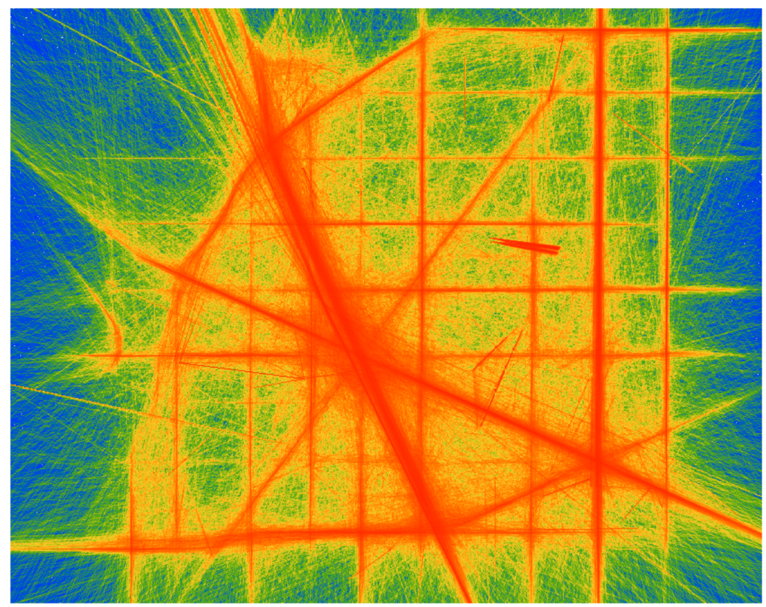

LBS data trips

More than 300,000 individuals



Median trip duration 28 mins

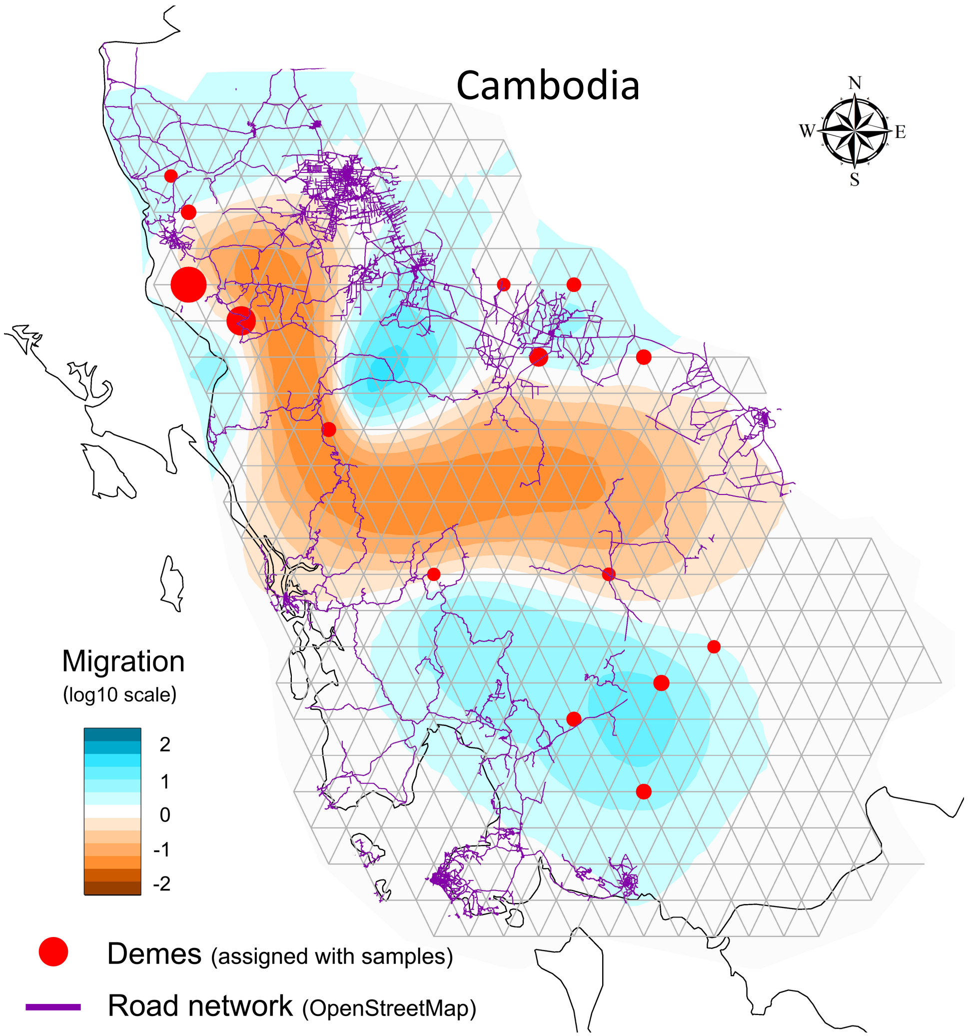
Median trip distance 1.8 mile



Median trip duration 2 mins

Median trip distance 0.2 mile

# Other new sources of data, e.g., genomic data also bring important opportunities to understanding mobility in new ways

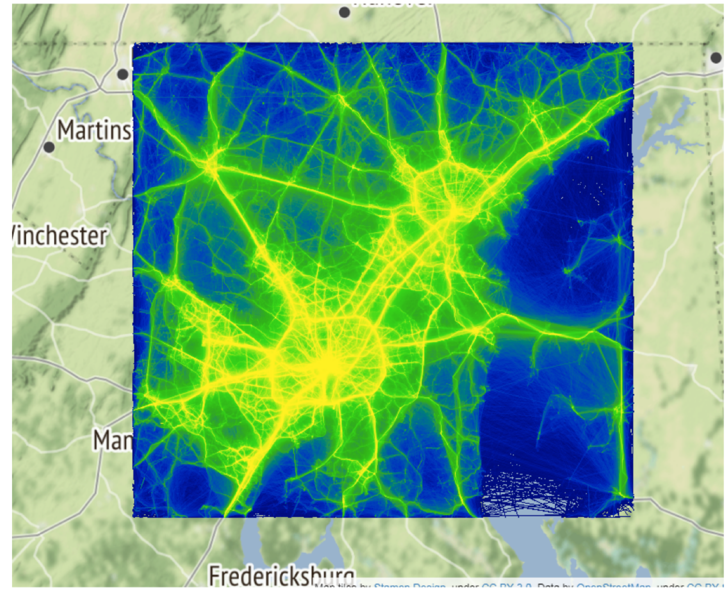



**Malaria elimination in the Greater Mekong Subregion**

Use big data techniques to simulate and estimate patterns of **parasite gene flow** with a **human mobility travel network**

Y. Li

# Contact:




**Kathleen Stewart, PhD**

**Director, Center for Geospatial Information Science**

**Professor, Department of Geographical Sciences**

**stewartk@umd.edu**

**Tel: 301-405-3203**

**https://geospatial.umd.edu**

This research is supported in part by grants from FHWA (693JJ31750013 and 590523) and NIAID-NIH (U19AI129386).

Research shown involves: J. Fan, Y. Li, Maryland Transportation Institute, and Center for Vaccine Development and Global Health UMB