Tracking Human Mobility & Behavior during Pandemics: A GPS-Based Analysis of Movement Patterns & Social Distancing

College of Public Health
DEPARTMENT OF HEALTH ADMINISTRATION AND POLICY
George Mason University

Naren Durbha, MSHI; Janusz Wojtusiak, PhD

Health Informatics Program, College of Public Health, George Mason University

1. Introduction

- Analyzing movement behaviors is key in controlling the spread of infectious diseases
- Traditional aggregated data limits analysis, highlighting the need for individual-level data integration
- Aims to analyze and understand movement patterns at individual level during COVID19
- Data was collected in 2 phases from George Mason University participants

2. Data Collection - Methods

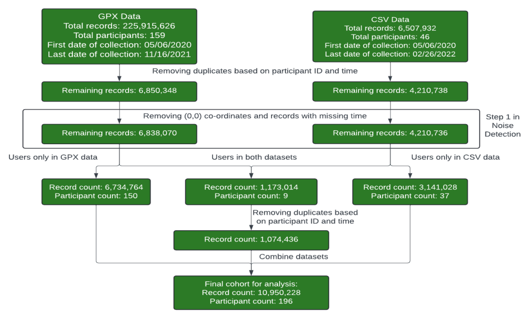
- In Stage 1, participants shared GPS data via mobile apps (iPhone/Android) and submitted daily symptom reports through Mason HealthCheck
- In Stage 2, WiFi connection logs were retrieved from university IT services, providing time and location data on campus
- Vaccination status and a behavioral survey with over 50 questions were collected to further understand participants' behaviors

3. GPS Data Processing & Analysis - Workflow & Results

Workflow Stage	*Methodology*	*Key Findings/Output*

I. Raw GPS Data Cleaning

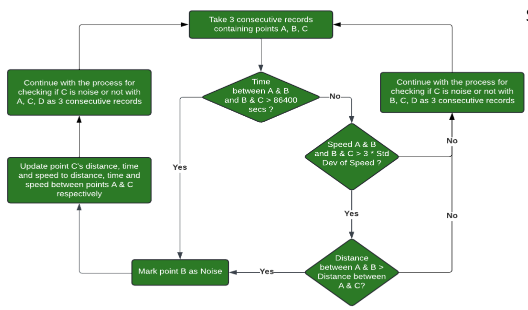
- Data were received in 2 formats: GPX and CSV
- iOS sent GPX in most cases, while most Android uploads were CSV



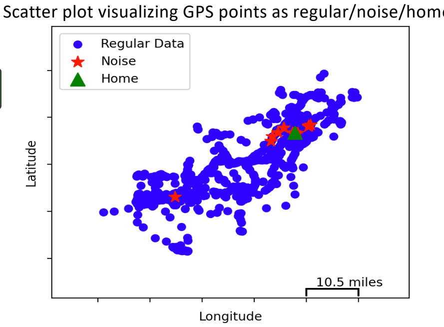
- Although 376 participants signed up for the study, due to discontinued participation or non-compliance, GPS data is available for 196 participants only
- Final cohort was used for noise & home detection per participant

II. Noise & Home Detection

- Designated home as the location most consistently observed between 10PM and 5AM



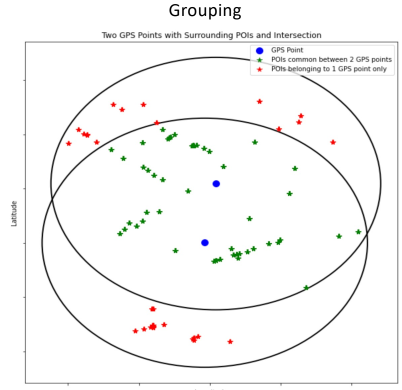
Scatter plot visualizing GPS points as regular/noise/home



III. Integration with SafeGraph Data & Consecutive Visit Grouping

- Matched cleaned GPS points to candidate POIs using a radius sweep (10 – 500 m)
- Collapse temporally adjacent points tied to same POI into a single visit

Scatter plot depicting distribution Consecutive Visit Grouping



- **Scale:** 232,423,558 raw GPS points → 10,950,228 after preprocessing → 10,248,410 after final filters; 6,638,890 unique locations
- **POI matching:** cumulative scores ranged 0–20,000; 68% of points mapped to a single POI candidate; 8% had 2–5 candidates; the rest had >5 (lower confidence)

- For each visit group, compute a cumulative score that increases with time spent and with log-distance from the POI
- Assign visit to the POI with the lowest mean cumulative score after applying global ambiguity filters
- Healthcare, recreation & food services were identified as frequently visited locations

4. Conclusion

- Data integration allowed analysis of frequent destinations and policy adherence, offering a framework to improve public health strategies