Location-Based Data Privacy Solution

Professors Matthew Schneider & Jordan L. Fischer, Esq.
Cameron Bale, PhD Student

Team: Data Privacy Experts in Statistics and Law

Global Legal Hackathon
May 22, 2020

Disclaimer: These materials do not constitute legal advice. The speakers do not warrant that the presentations or materials are free of errors, or will continue to be accurate. Opinions expressed are those of the speakers and statements in the presentations and the materials should be verified before relying on them.

Project Goal:

Location-based Data Privacy Solution

How do you trade-off the accuracy of location data with privacy?

⟶ Matching legally defined terms to use cases of location data.

⟶ Statistical solution to reduce the identifiability of individuals while maintaining most of the usefulness of the data.

Drexel

Privacy Legal Frameworks

Anonymization:

A process that removes the association between the identifying dataset and the data subject

Pseudonymization:

The processing of personal information in a manner that renders the personal information no longer attributable to a specific consumer without the use of additional information

De-Identification:

Information that cannot reasonably identify, relate to, describe, be capable of being associated with, or be linked, directly or indirectly, to a particular individual

Aggregation:

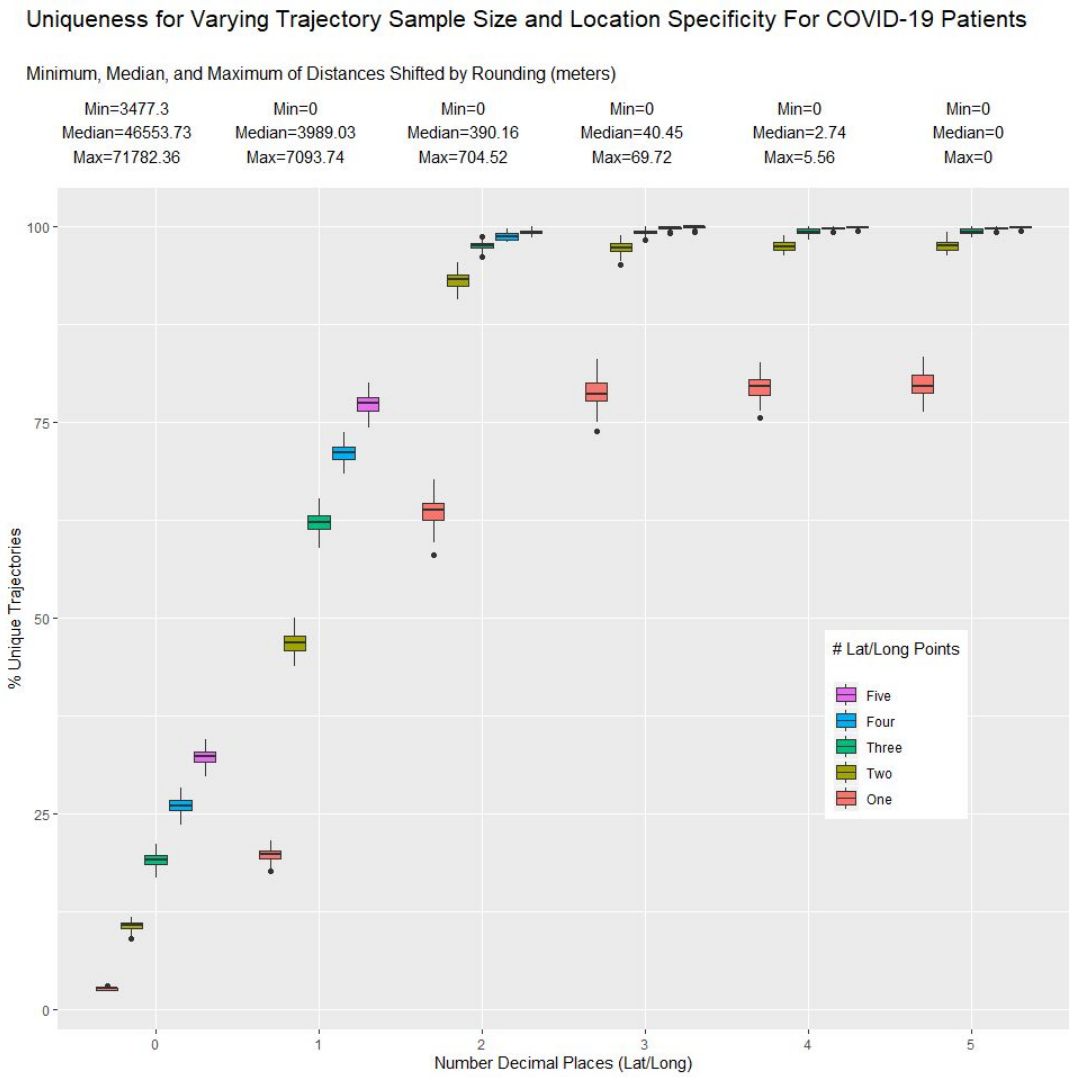
Information that relates to a group or category of individuals, from which individual identities have been removed, that is not linked or reasonably linkable to any individual

Statistical Analysis:

→ Location coursening via rounding:
round latitude/longitude values
to reduce location granularity

→ Improves privacy in two ways:
1. Reduces location uniqueness
2. Alters original locations by
a controllable distance

https://github.com/cdbale/Hackathon



Metric #1: Neighborhood Risk

⟶ The number of COVID-19 persons visiting neighborhood **k** on day **t**

$$Prevalence_{t,k} = \frac{\sum_{i=1}^{i=n} I(GPS_{i,t} \in Neighborhood_k)}{Population_k}$$

⟶ Legal Term: Aggregation (HIGH)

⟶ Examples: South Korea

⟶ Privacy Solution: Location coarsening

Differential privacy-based histogram with noise infusion

Drexel

Metric #2: Specific Location Risk

⟶ The number of COVID-19 persons at specific location **k** on day **t**

$$Threat_{t,k} = \sum_{i=1}^{i=n} I(GPS_{i,t} \in GeoFence_k)$$

⟶ Legal Term: Aggregation and De-Identification (MEDIUM)

⟶ Examples: South Korea

⟶ Privacy Solution: Location coarsening

Drexel

Metric #3: Contact Tracing

⟶ Did person **i** have contact with person **j** at any time **t** within distance **d** ?

$$If \left\| GPS_{i,t} - GPS_{j,t} \right\| < d \; then \; contact_{i,j} = 1$$

... results in a "social network" data like Facebook

⟶ Legal Term: Anonymization and Pseudonymisation (LOW)

⟶ Examples: South Korea, Israel, Singapore, many others (Bluetooth and GPS)

⟶ Privacy Solution: Privacy-preserving graph degree sequence but fundamental issues remain

Drexel

Metric #4: Quarantine Compliance

→ Did person **i** leave location **k** at any time **t**?

$$If\ \left(GPS_{i,t} \notin GeoFence_k\right)\ then\ Violation_i = 1$$

→ Legal Term: None (NON-EXISTENT)

→ Examples: Hong Kong, Poland

→ Privacy Solution: Access Control

Drexel

Future Work:

⟶ Identifying and Defining the Privacy Risks

⟶ Further define key legal terms & factors with corresponding statistical analysis

⟶ Differentially private histogram approach for geospatial data which provides a theoretical guarantee to individual privacy

Drexel

Contact Us:

Matthew Schneider
Assistant Professor of Statistics
mjs624@drexel.edu

Cameron Bale
PhD Student in Business Analytics
LeBow College of Business
cdb327@drexel.edu

Jordan L. Fischer, Esq.
Teaching Professor of Law
Thomas R. Kline School of Law
jlf324@drexel.edu
267-536-9376

Disclaimer: These materials do not constitute legal advice. The speakers do not warrant that the presentations or materials are free of errors, or will continue to be accurate. Opinions expressed are those of the speakers and statements in the presentations and the materials should be verified before relying on them.