Understanding Mobility in Sierra Leone During Covid-19 Using Call Detail Records

Thesis Defense

Yanchao Li
June 2021

Background

Sierra Leone
a developing country located on the southwest coast of West Africa

Research Area
Western Area Urban & Western Area Rural

Covid19:

March 30th -- first case of Covid-19

Three day lockdown & 14 day inter-district travel ban

Using call detail records (CDR) data to understand the effectiveness of policy actions

Research Area

Research Question

How did travel policies and socio-economic status impact mobile user’s accessibility to services during the COVID-19?

Why is it important?

Literature

Most literature used aggregated metrics (e.g. OD flows) to understand the risk of Covid-19, but we focus on the individual level mobility patterns change and combine with socioeconomic data to define user typologies

Combine different metrics together and build a big data processing pipeline with Spark. Examine whether methods in previous literature are workable in the Sierra Leone context and apply different thresholds

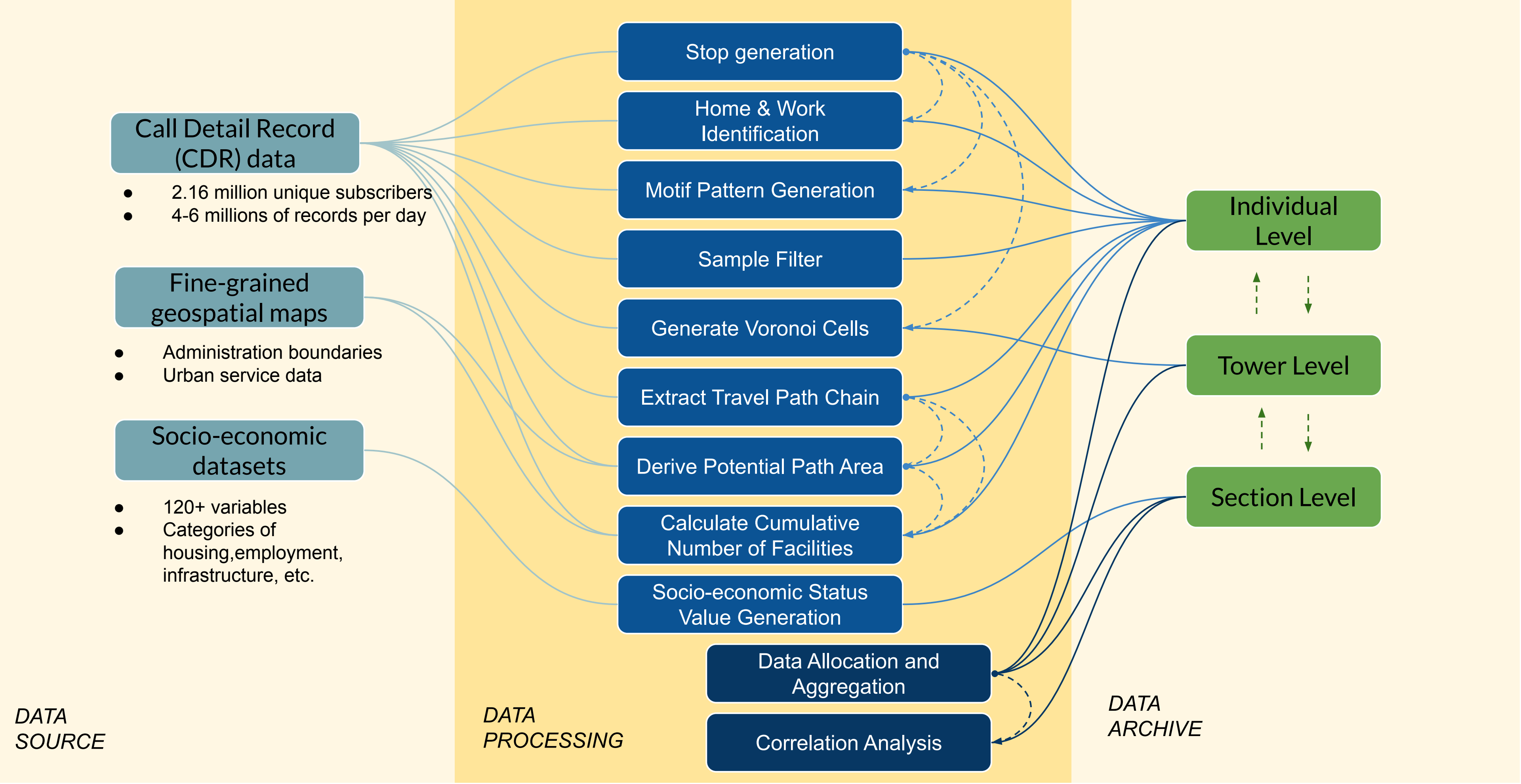
Practice

Aid policy makers in understanding public compliance to their Covid-19 mitigation measures

Inform the government of the vulnerable groups of people under the impact of travel restriction policies

Help the government for future decisions on travel bans.

Method & Data

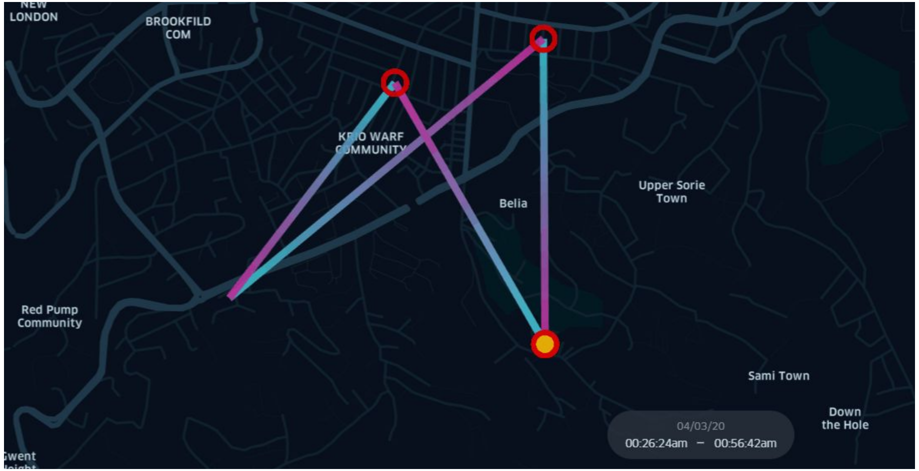

Call Detail Record (CDR) data
2.16 million unique subscribers
4-6 millions of records per day
Fine-grained geospatial maps
Administration boundaries
Urban service data
Socio-economic datasets
120+ variables
Categories of housing,employment, infrastructure, etc.
DATA SOURCE
Stop generation
Home & Work Identification
Motif Pattern Generation
Sample Filter
Generate Voronoi Cells
Extract Travel Path Chain
Derive Potential Path Area
Calculate Cumulative Number of Facilities
Socio-economic Status Value Generation
Data Allocation and Aggregation
Correlation Analysis
DATA PROCESSING
Individual Level
Tower Level
Section Level
DATA ARCHIVE

Stop Generation & Home/ Work Location Identification

Valid Stop: with at least two continuous events and event-gap should be between 10 minutes and 4 hours.

Home: user has the highest stop frequency during daily nighttime hours (8 PM to 8 AM from Monday to Sunday)

Work: user has the highest stop frequency during normal business hours (9 AM to 5 PM from Monday to Friday)



Raw CDR Data Visualization

Attributes	Description	Sample
Antenna_ID	Tower index of either incoming or outgoing calls or SMS	11024
Lat	Tower location (Latitude)	-13.218795
Long	Tower location (Longitude)	8.480993
Direction	Incoming / Outgoing / SMS	Incoming
Source	Hash of the phone number of the subscriber logging the event (either incoming or outcoming call) CDR record	00029E72B5BDC783 E696E4827A0776E5
Target	Hash of the phone number of the subscriber logging the event (either incoming or outcoming call) CDR record	FFD4233B0637F668F 3A1CCD89B7B0274
Duration	Length of the call in seconds	1760
Timestamp	Datetime marking the the beginning of the call or sms	2020-03-04 20:00:10

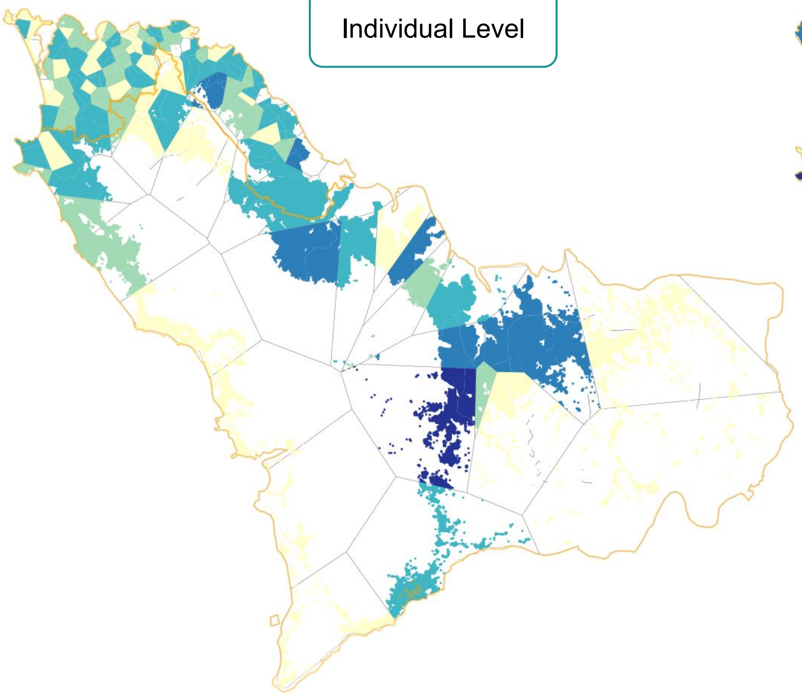
Raw CDR Data Schema

Data Allocation and Aggregation

1) Generate voronoi cells and allocate users to the corresponding voronoi cell;
2) Iterate voronoi cells. For each voronoi cell, allocate users to different land use types based on varied weights;
3) Iterate administrative sections. For each section,sum up the users belonging to the section;
4) Calculate the expectation of users' accessibility / mobility index in the administrative section level.

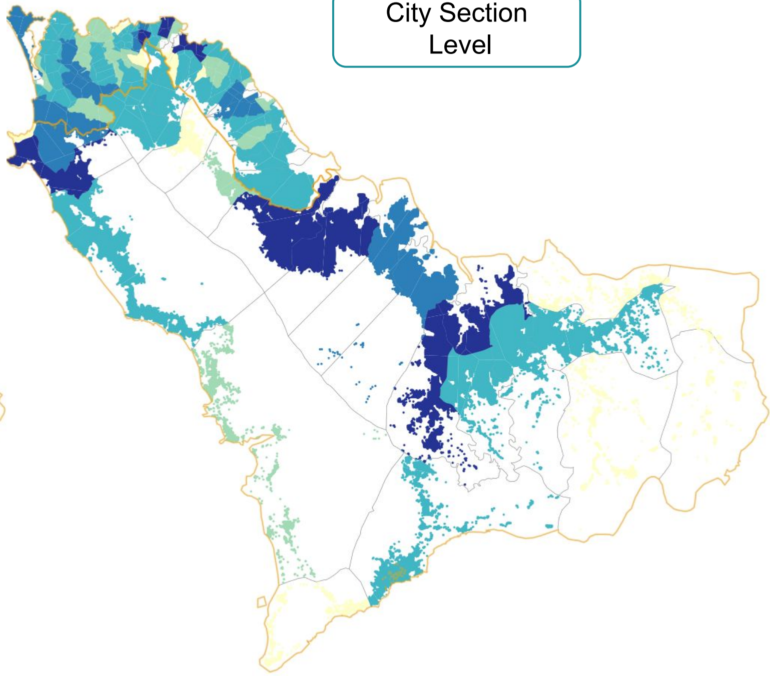
Number of users in each voronoi cell

Individual Level



Number of users in each section

City Section Level



$$u_s = \sum_{k=1}^{k} n_k$$

$$n_k = \frac{\sum_{l=1}^{l} s_{kl} \times b_l \times p_q}{\sum_{l=1}^{l} s_{ql} \times b_l}$$

n is the number of users allocated to a shard;p_q is the users number of a voronoi cells q; s_l is the area of a type of land use, e.g. commercial area, water body; b_l is the coefficient of users weight to the type of land use.

Number of users

- <500
- 400 ~ 1000
- 1000 ~ 2000
- 2000 ~ 3000
- > 3000

Socio-economic Status Value (SES) Generation

1) Run descriptive analysis to normalize all 123 variables
2) Calculate SES via Principal Component Analysis (used top five components to represent the variance of the dataset)
3) Group administrative sections into six SES categories using equal-interval classification

$$SES = \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \ldots + \lambda_t Q_t ,$$

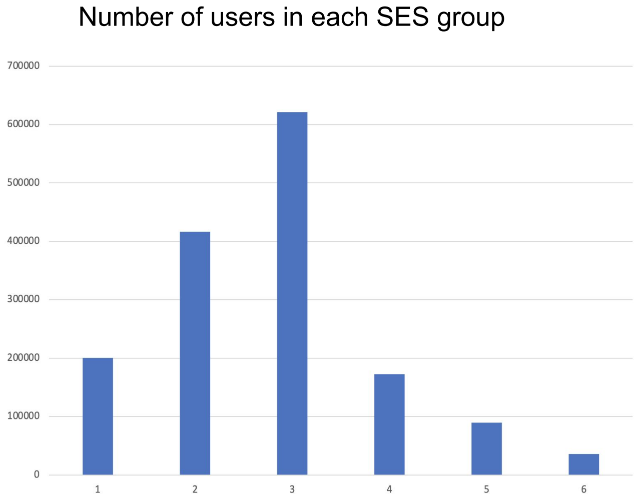
$$Q_t = S_{t1}X_1 + S_{t2}X_2 + S_{t3}X_3 + \ldots + S_{tp}X_p ,$$

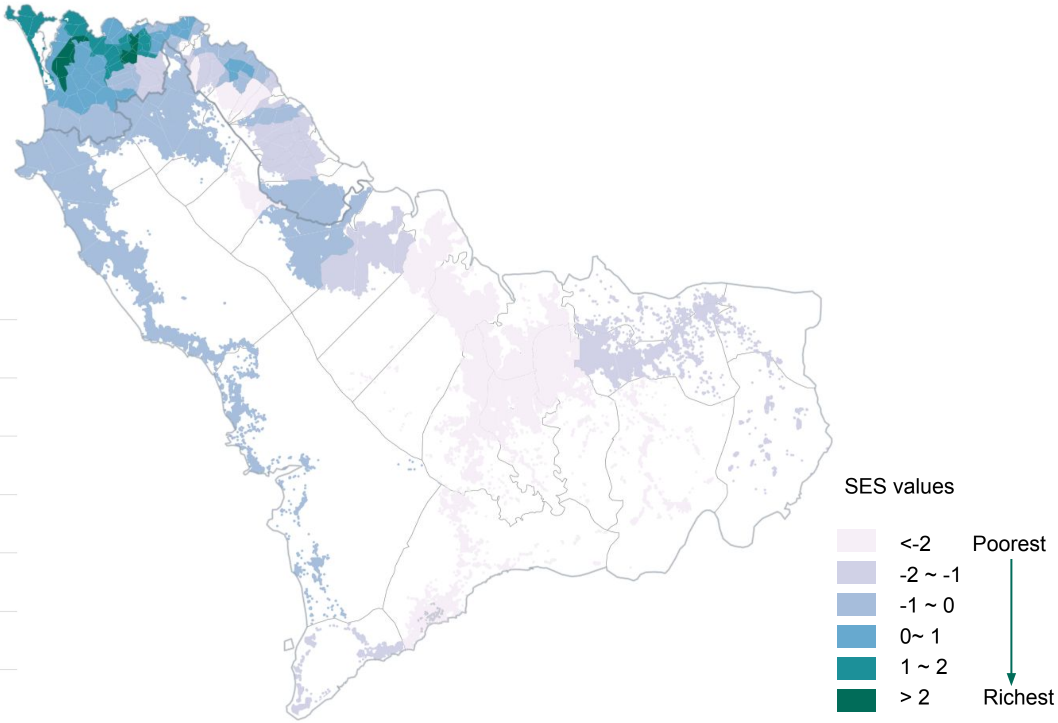
λ_t are the eigenvalues corresponding to the principal components;

p is the number of variables,

S_{tp} are the contribution factors corresponding to the principal components,

and X_p are the standardized values of the census dataset.





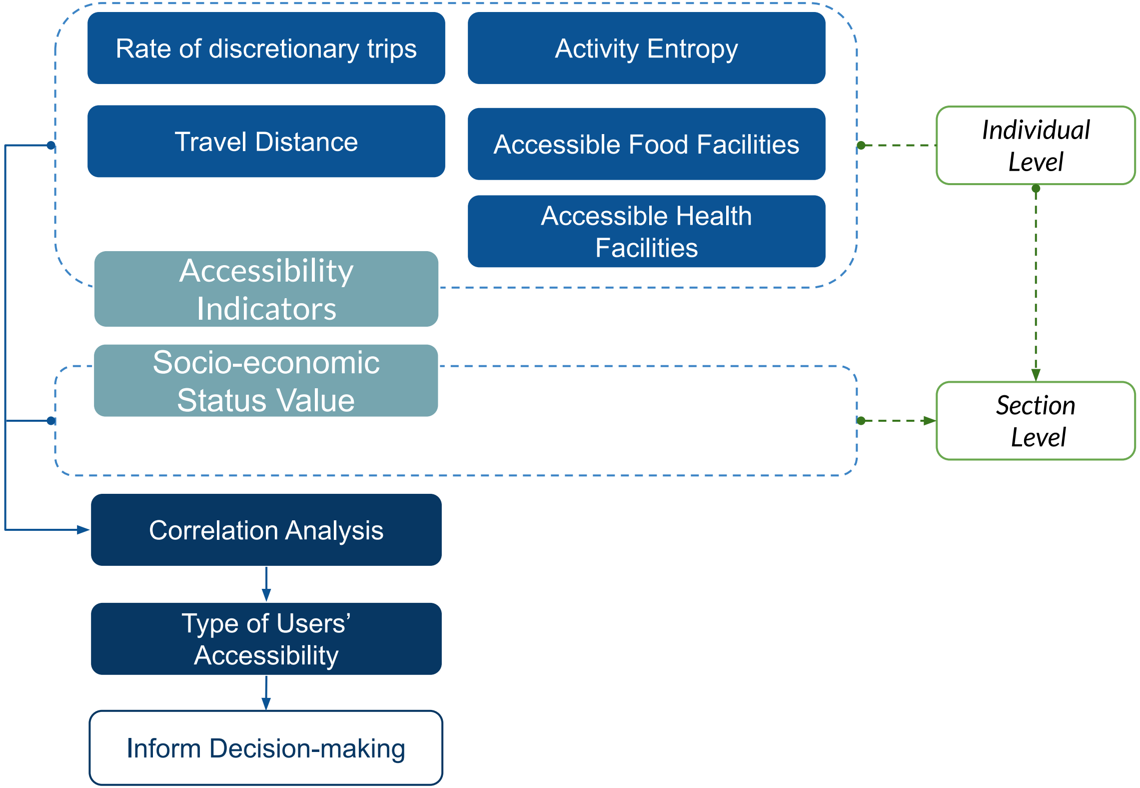
How do travel policies and socio-economic status impact mobile user's accessibility to services during the COVID-19?

(1) How to measure users' accessibility to services/ opportunities (i.e. food, healthcares)?

(2) How does users' accessibility change in the context of travel restriction policies?

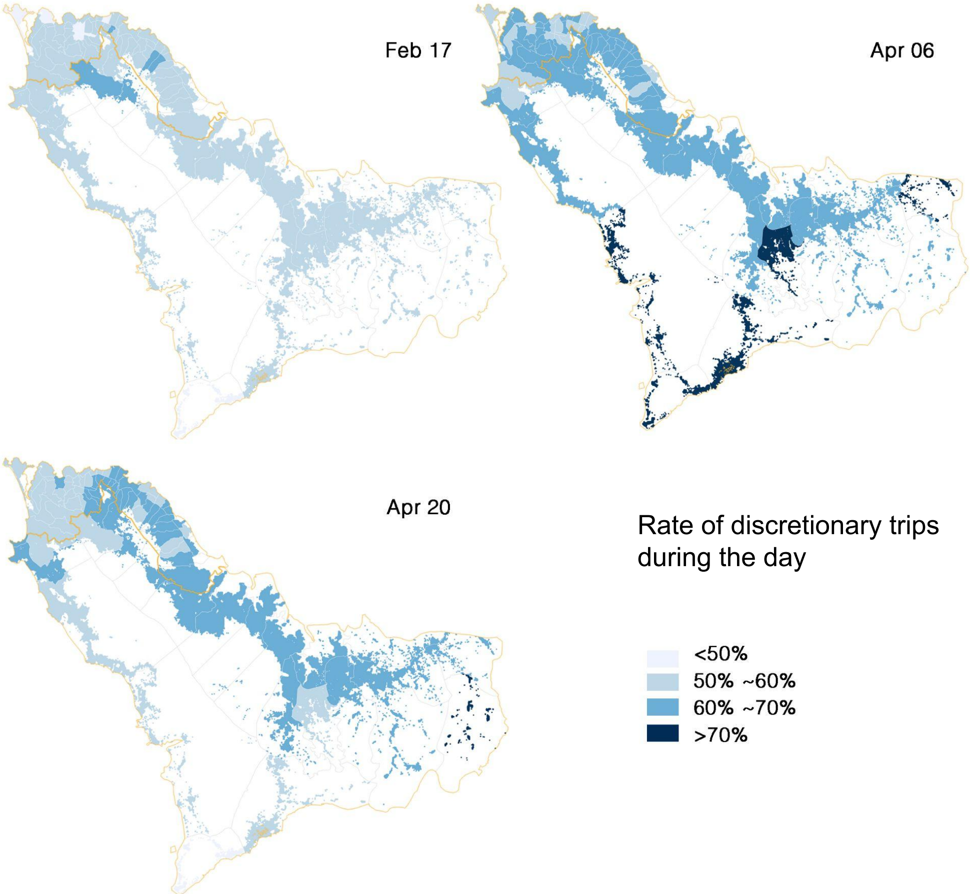
(3) How does the change of accessibility relate to users' socioeconomic status?

And (4), how to improve accessibility to have helpful travel restriction policies?

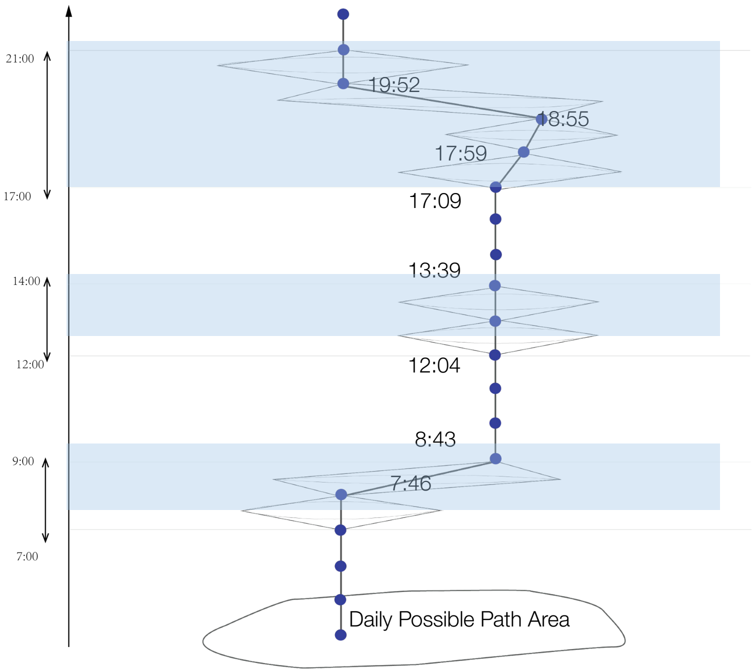


Loose-travel-ban and lock-down policies effectively changed people’s accessibility behaviors.

1) Discretionary Trips Increased under the Pandemic.



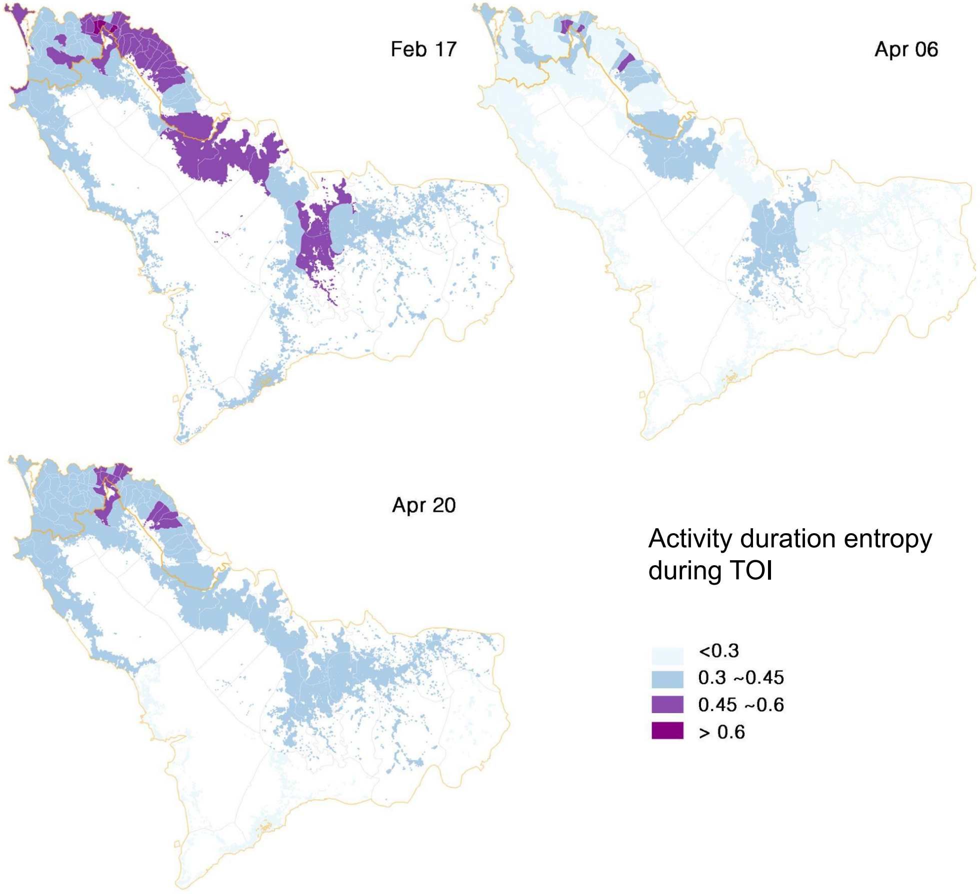
Extract chronologically ordered path chain during **time period of interest (TOI)**
{STP7, STP8, STP13, STP14, STP17, STP18, STP19, STP20}



mobile user ID:
AC7D5369B7CD4517DDA9E3BAC5C6

Loose-travel-ban and lock-down policies effectively changed people’s accessibility behaviors.

2) Travel Plans Were Simplified under the Pandemic .



Activity entropy

– focus on time duration at each location;

– measures the diversity of users’ travels;

– a large value of this indicator suggests that the diversity of a user's daily activities is high, a small value suggests that the activity diversity is low.

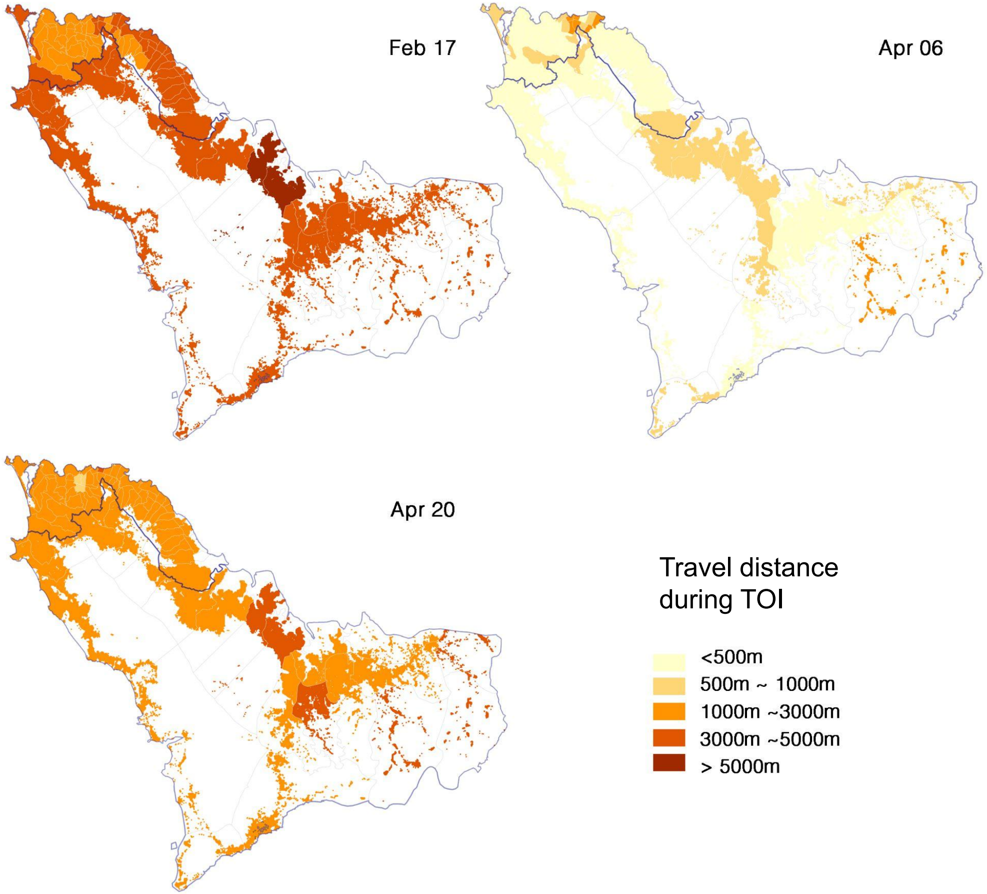
$$h_j = - \sum_{k=1}^{k} p_k log(p_k)$$

Where $\sum_{k=1}^{k} p_k = 1$

p is the proportion of duration of staying at a location *k*.

Loose-travel-ban and lock-down policies effectively changed people's accessibility behaviors.

3) Travel Distance for Discretionary Activities Sharply Decreased .



Travel distance

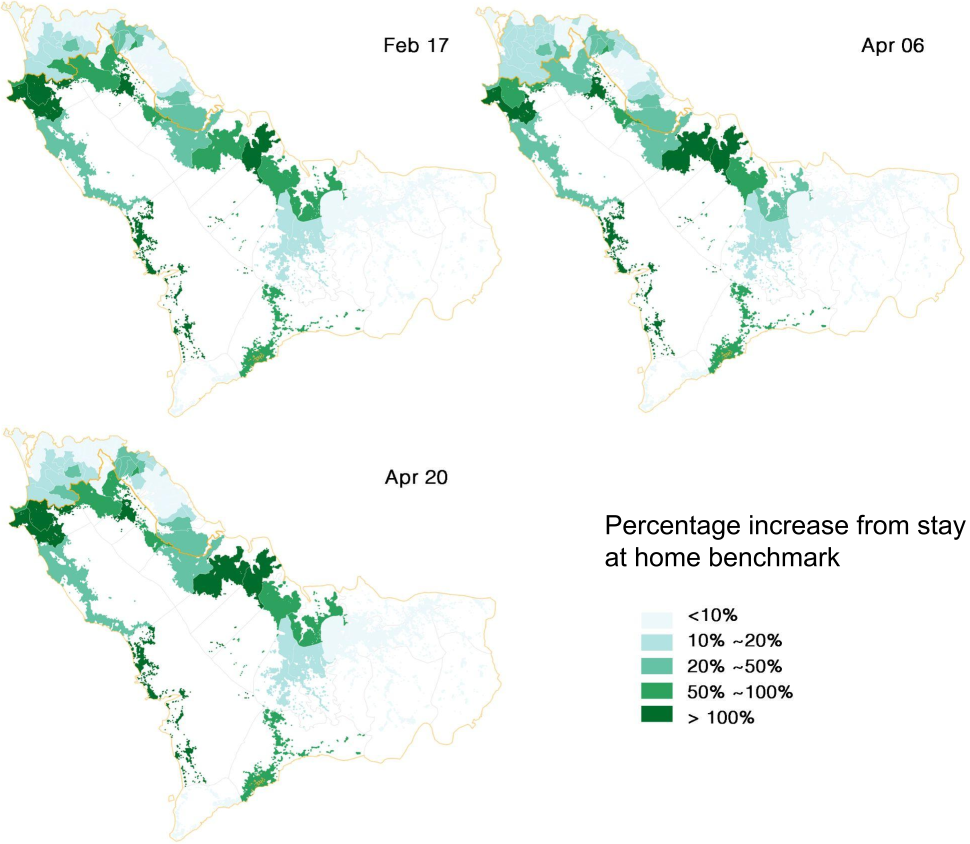
For each user, the travel distance *D* for discretionary activities is calculated as:

$$D = \sum_{i=1}^{n} (X_{i+1} - X_i) \quad (n \subseteq R,\ X \subseteq R^2)$$

X is a vector representing for the anchor point location

For users in most of the city sections, loose-travel-ban and locked-down policies seem to have no obvious impact on their daily available food and healthcares.

1) The Rate of Available Increased Food Resources Compared with the Stay-at-home Benchmark Does Not Change Obviously.



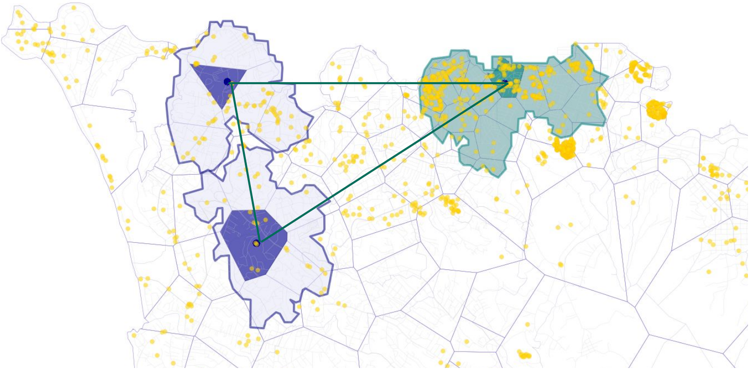
Increased Food Resource

$$CUM_j = \sum_{k=1}^{n} \sum_{f}^{ALL} \delta_f$$

$$\delta_f = \{1 \; if \; f \; belongs \; to \; potential \; path \; area; \; or \; 0 \; otherwise\}$$

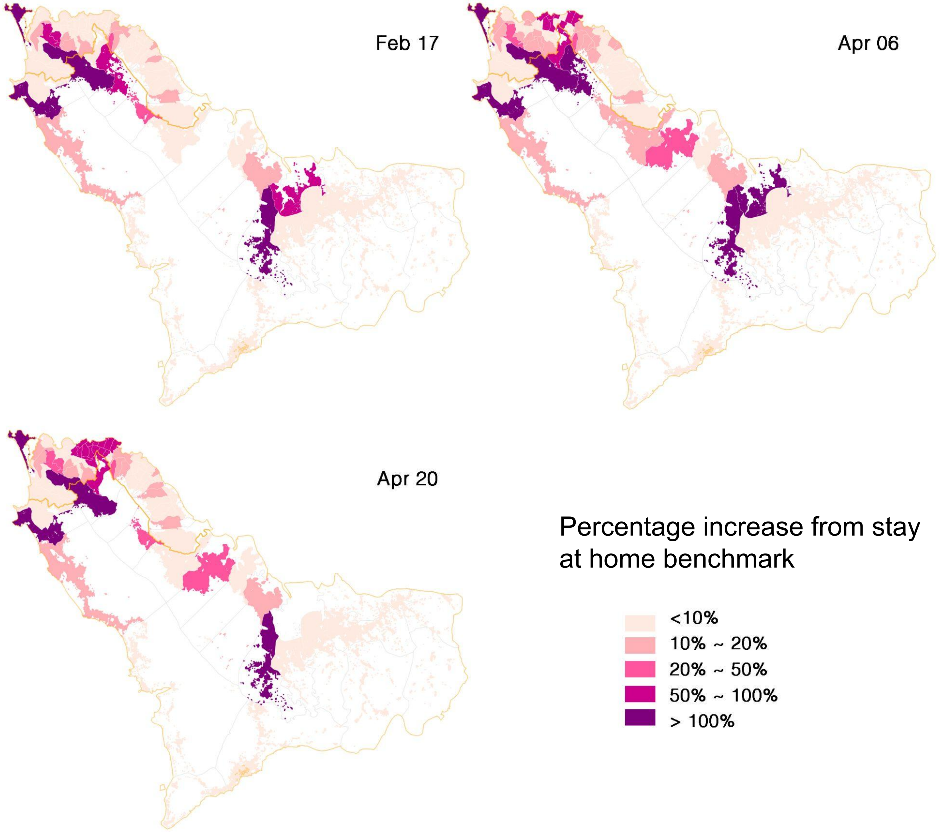
$$increase \; rate \; r = \frac{CUM_{experimental} - CUM_{control}}{CUM_{control}} \times 100\%$$

Potential path area: the 20-min walking area generated based on fine-grained road network.



For users in most of the city sections, loose-travel-ban and locked-down policies seem to have no obvious impact on their daily available food and healthcares.

2) The Rate of Available Increased Health Resources Compared with the Stay-at-home Benchmark Does Not Change Obviously.



Increased Health Resource

$$CUM_j = \sum_{k=1}^{n} \sum_{f}^{ALL} \delta_f$$

$$\delta_f = \{1 \textit{ if } f \textit{ belongs to potential path area}; \textit{ or } 0 \textit{ otherwise}\}$$

$$\textit{increase rate } r = \frac{CUM_{experimental} - CUM_{control}}{CUM_{control}} \times 100\%$$

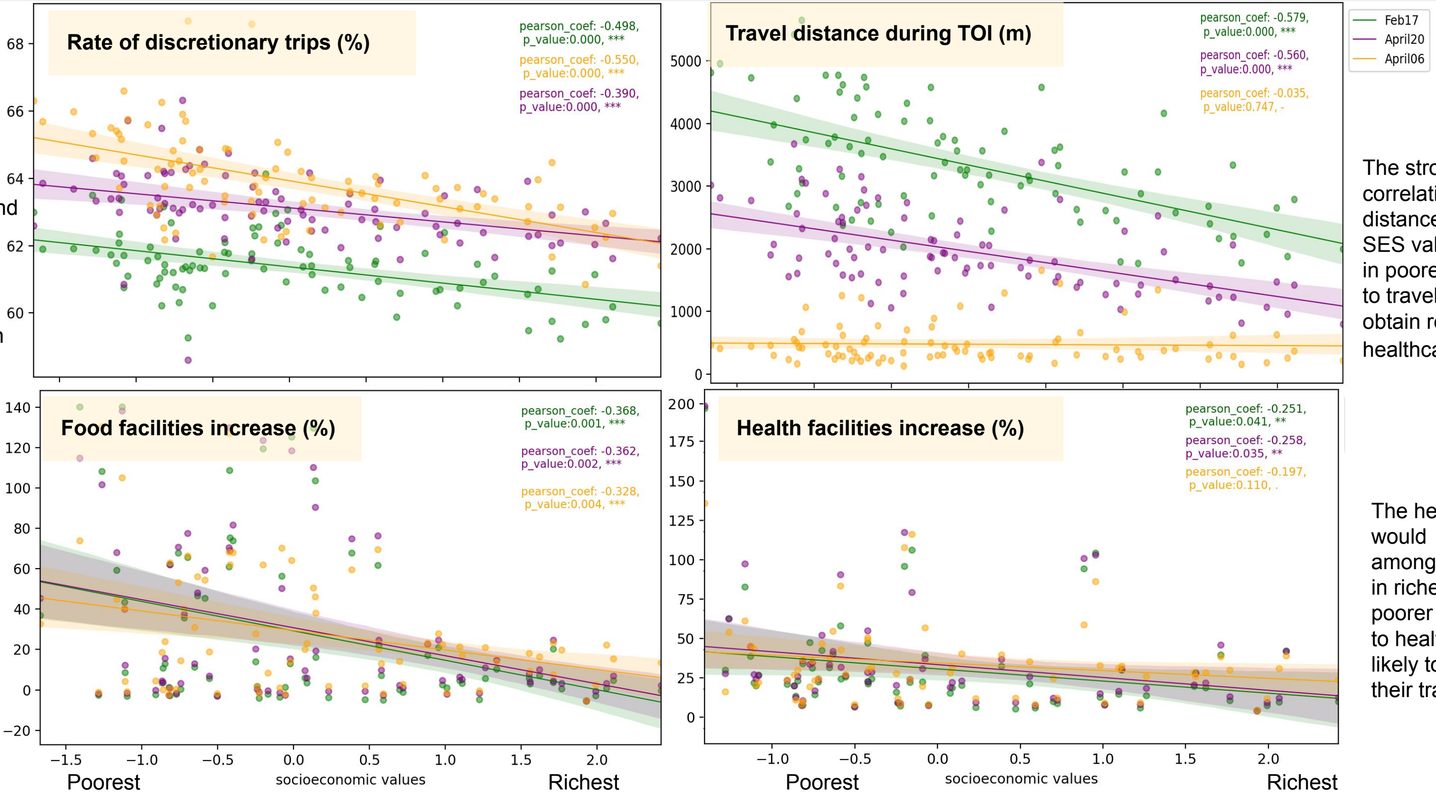
Potential path area: the 20-min walking area generated based on fine-grained road network.

Users of different socioeconomic status (SES) have different accessibility behaviors.

Users of Lower Socioeconomic Status:

- Have a longer travel distance for discretionary activities;
- Have a higher percentage of discretionary trips.
- Available food/ healthcare resources are more likely to be impacted by their travel behaviors.

Richer users tend to have lower rate of discretionary trips, while poorer users tend to have higher; which indicates people with lower socioeconomic status tend to travel more for discretionary activities such as obtaining food.



The strong negative correlation between travel distance during TOI and SES values suggests users in poorer city sections tend to travel longer distance to obtain resources (i.e., foods, healthcares, etc.)

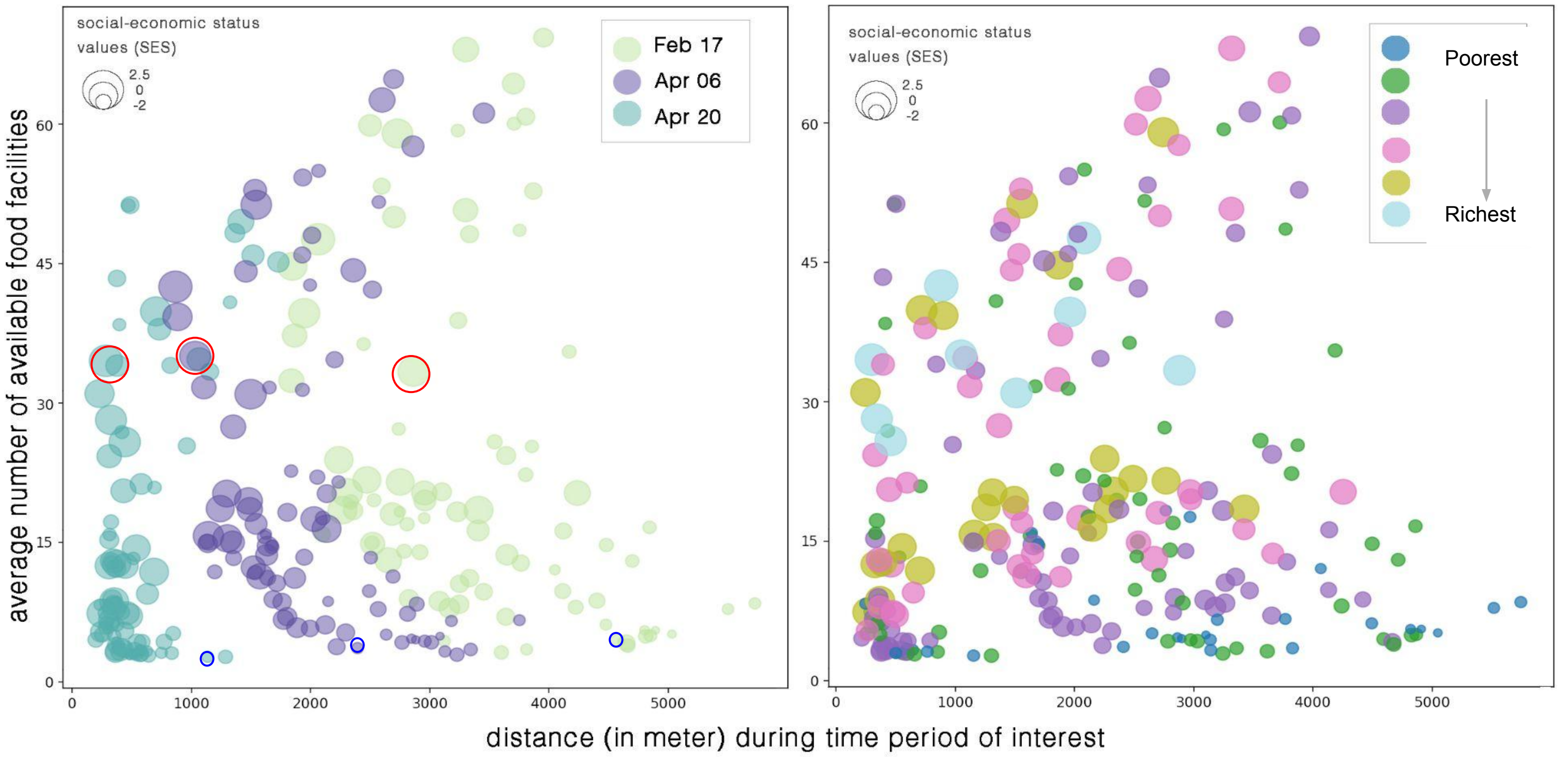
If facilities don't increase, that means users' accessibility to foods would be less likely to be impacted by their travel patterns.

The health resources would increase more among poorer regions than in richer regions, indicating poorer users' accessibility to healthcares would more likely to be impacted by their travel patterns.

Users of different socioeconomic status (SES) have different accessibility behaviors.

Not All Users' Available Food Facilities Increase When Their Travel Distance Increases.

Decomposition of travel distance and accessible food facilities



Users are in different levels of vulnerability regarding to accessibility to food and healthcares resources.

Type One:

Have higher SES, no obvious change in travel distance and accessibility to resources.

Type Two:

Have limited accessibility during the lockdown time, but can recover when long-distance travel are allowed.

Type Three:

Have enough availability to food/health resources, but available resources increased with a slightly increase of travel distance.

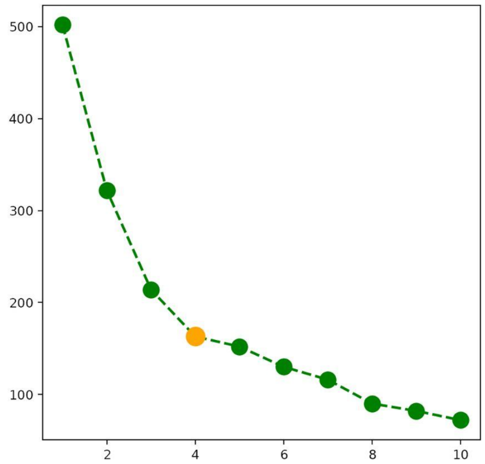
Type Four:

Have limited accessibility both in the pre-pandemic and lock-down period.

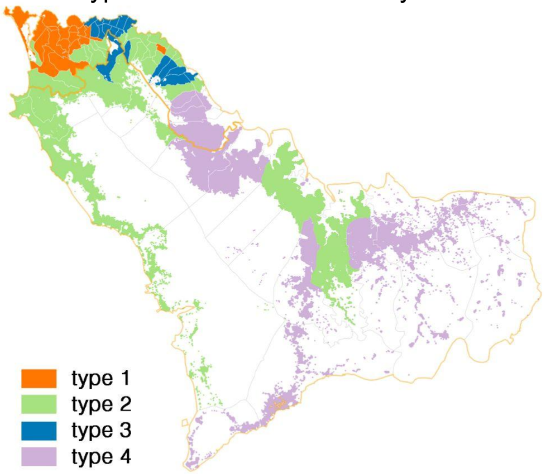
Within cluster sum of the squares was used to determine the optimal K.

It shows that K = 4 is the elbow point.

Optimal K choosing



Type of users in accessibility to resources



Travel restriction policies effectively changed people's accessibility behaviors.

During the pandemic period: the rate of discretionary activities increased; travel distance for discretionary activities sharply decreased; travel plans were simplified.

For users in most of the city sections, travel restriction policies seem to have no obvious impact on their daily available food and healthcares.

Users of different socioeconomic status (SES) have different accessibility behaviors.

Users of lower socioeconomic status:
have a longer travel distance for discretionary activities; have a higher percentage of trip plans for obtaining food/healthcare resources.Available food/ healthcare resources are more likely to be impacted by their travel behaviors.

Users are in different levels of vulnerability regarding to accessibility to food and healthcares resources.

Summary:
impact of travel policy and SES values to accessibility indicators

Indicators	**Travel policies**	**Socio-economic status values**
	Compare lockdown period with pre-pandemic period	Positive impact (P) Negative impact (N) No obvious impact (–)
Percentage of discretionary trips during TOI divided by total trips	Increased	N
Travel distance during TOI	Decreased	N
Entropy of travel time duration during TOI	Decreased	–
Increase of food accessibility compared with staying-home benchmark	No obvious change	N
Increase of food accessibility compared with staying-home benchmark	No obvious change	N

Limitations

- Use tower approximation instead of allocating the trips to the actual roadmap;
- Limitation of sparsity for infer actual movements;
- Informal food resources (e.g. street vendors) were not completely included in our research;
- Imputing missing values in the datasets.

Future Research

Other factors may influence accessibility/ mobility patterns are :

- Former Ebola outbreak,
- Local policy and regulation,
- and other socio-economic factors e.g., education level;
- ...

Thanks!

Defense Questions