



UK-South Korea Geospatial Data Science Exchange Seminar

Exploratory Analysis of Smartphone-Based VGI Datasets Collected by Drivers in the Daegu Metropolitan City Area: Focusing on the Role of Time Geography in Data Pre-processing

June 9,
2023

Presenter: Jae Hyun Lee¹

Contributors: Soojeong Choi,¹Jiyeon Shim,¹Sanghyeon Yun,²Seo Youn Yoon(KRIHS)³, Myeongho Yeo(WeDrive)⁴

¹ GIScience & Geospatial Bigdata Laboratory
Department of geography / Kyungpook National University

² Climate Science Laboratory
Department of geography / Kyungpook National University

³ Korea Research Institute for Human Settlements

⁴ WeDrive

Outline

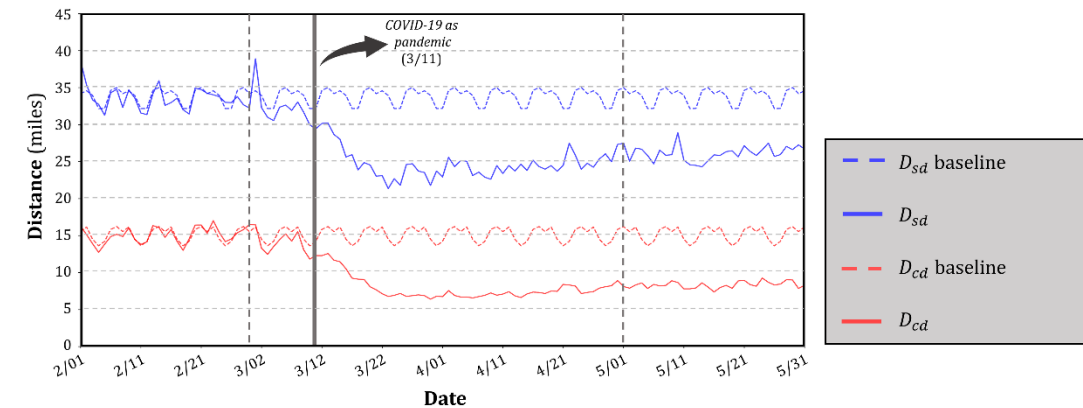
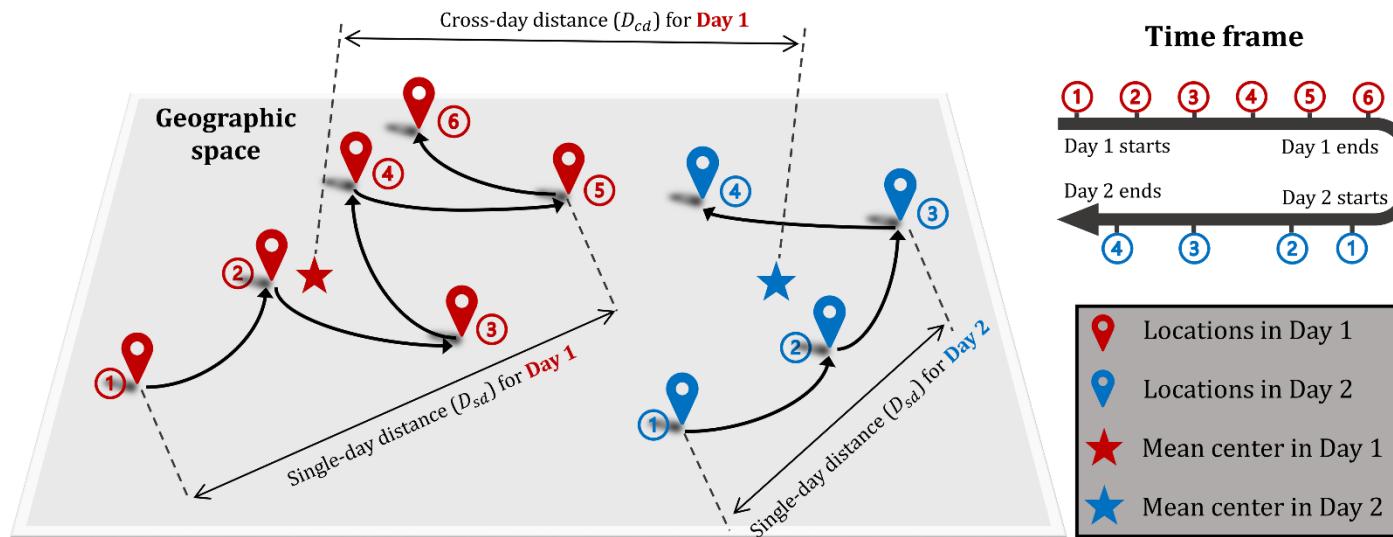
1. Introduction
2. Data
3. Case study & Potential Applications
4. Summary & Discussion

1.

Introduction

Introduction

- Though many urban transportation models are developed with an assumption of a “**typical day**”, in fact, **every day is different** in terms of the interactions between urban environments and residents. Also, the interactions can vary within a day, day-to-day, weeks, seasons, and years
- During **major disasters or outbreaks** like the recent pandemic, **there is no “typical day”**. Therefore, we need to have new data and methods to understand this dynamics of mobility(travel behavior).



Sources: Huang et al. (2020)

Introduction

| Three levels of Travel Behavior(a.k.a. Human mobility) Dynamics

1) Microdynamics

- Concerned with the time of day activity-scheduling behaviors, changes in accessibilities, travel demands that in 24 hours

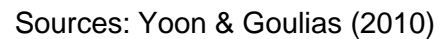
2) Macrodynamics

- Deals with the life-span of individuals, such as the progression in the life course(birth / death / change in life cycle stages)

3) Mesodynamics

- Travel behavior with some regularity that are located between daily dynamics and long-term dynamics

Microdynamics: Time of day dynamics



<Sample Data - Actively provided(every 5 years)>

Introduction

Macrodynamics: Changes in behavior along the life courses



PLACE 1 BEGIN RECORDING YOUR TRAVEL HERE

Your travel day begins with Place 1 at 3 a.m. Most people are home asleep at this time. If this is the case with you, check "My Home," then record your activities at this place and the exact time you leave.

REMEMBER!

PLACE 1 BEGIN RECORDING YOUR TRAVEL HERE

Your travel day begins with Place 1 at 3 a.m. Most people are home asleep at this time. If this is the case with you, check "My Home," then record your activities at this place and the exact time you leave.

PLACE 1 BEGIN RECORDING YOUR TRAVEL HERE

Your travel day begins with Place 1 at 3 a.m. Most people are home asleep at this time. If this is the case with you, check "My Home," then record your activities at this place and the exact time you leave.

REMEMBER!

A PLACE can be a transfer point such as a transit stop or a parking location.

WHERE were you at 3 a.m. on your travel day?

☐ My Home
☐ My Primary Job
☐ My School
☐ My Second Job
☐ Transit Stop (Bus or rail)
☐ Other Place

What ACTIVITIES did you do there? (Write up to 3 codes from LIST 2 on the flap of the back cover and the start and end times.)

Code: Specify if code "3B":

Activity 1: Start: : am / pm End: : am / pm

Activity 2: Start: : am / pm End: : am / pm

Activity 3: Start: : am / pm End: : am / pm

What TIME did you LEAVE? (Please record exact time)

: am / pm → Move on to next PLACE

☐ I didn't leave this place today → Answer D Below

What is the MAIN reason you didn't leave this place today?

☐ I was sick
☐ Vacation or personal day
☐ Child was sick
☐ Other household member was sick
☐ I am home-bound, elderly, or disabled
☐ Worked at home (for pay)

☐ Not scheduled to work
☐ Worked around home (not for pay)
☐ No transportation available
☐ Weather
☐ No reason to travel
☐ Other:

***If you didn't leave this place today and you did more than three activities, please record the codes and start and end times for each activity on a separate sheet of paper.**

Introduction

| Mesodynamics: Weekly and monthly dynamics?



Sources: Lee et al. (2016)

<Week-by-week Activity Space Growth Pattern>

PLACE 1

BEGIN RECORDING YOUR TRAVEL HERE

Your travel day begins with Place 1 at 3 a.m. Most people are home asleep at this time. If this is the case with you, check "My Home," then record your activities at this place and the exact time you leave.

REMEMBER!
A PLACE can be a transfer point such as a transit stop or a parking location.

A WHERE were you at 3 a.m. on your travel day?

☐ My Home
☐ My Primary Job
☐ My School
☐ My Second Job
☐ Transit Stop (bus or rail)
☐ Other Place

Name of Place: _____
Street Address or Nearest Cross-street: _____
City (Colonia): _____ State: _____ Zip Code: _____

B What ACTIVITIES did you do there? (Write up to 3 activities from LIST 2 on the flap of the back cover and the start and end times.)

Code: Specify if code "38": _____

Activity 1: _____ Start: _____ : _____ am / pm End: _____ : _____ am / pm
Activity 2: _____ Start: _____ : _____ am / pm End: _____ : _____ am / pm
Activity 3: _____ Start: _____ : _____ am / pm End: _____ : _____ am / pm

C What TIME did you LEAVE? (Please record exact time) _____ : _____ am / pm → **Move on to next PLACE**
☐ I didn't leave this place today → **Answer D Below**

D What is the MAIN reason you didn't leave this place today?

☐ I was sick
☐ Vacation or personal day
☐ Child was sick
☐ Other household member was sick
☐ I am home-bound, elderly, or disabled
☐ Worked at home (for pay)

☐ Not scheduled to work
☐ Worked around home (not for pay)
☐ No transportation available
☐ Weather
☐ No reason to travel
☐ Other: _____

**If you didn't leave this place today and you did more than three activities, please record the codes and start and end times for each activity on a separate sheet of paper.*

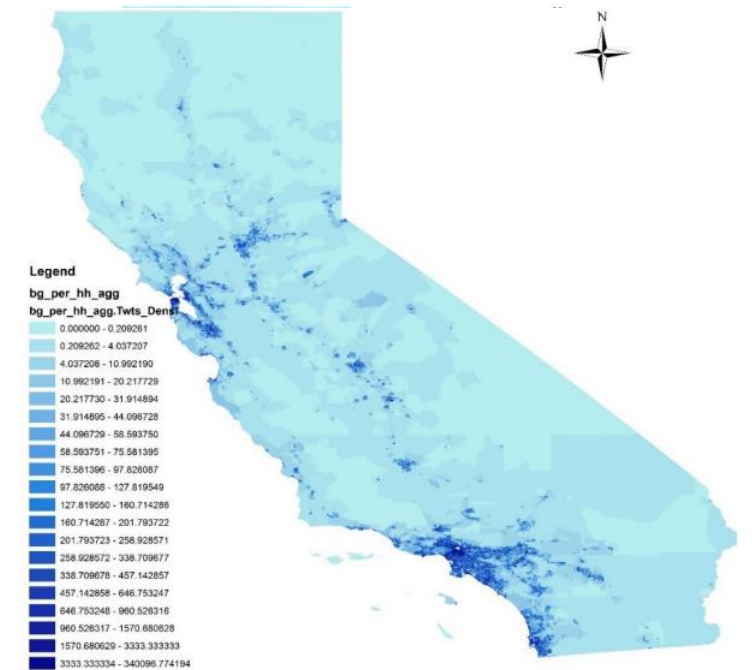
Introduction

| Mesodynamics: Weekly and monthly dynamics?



Sources: Lee et al. (2016)

<Week-by-week Activity Space Growth Pattern>



Sources: Lee (2016)

<Number of geo-tagged Tweets in Census Block Groups
(California 6 months, about 90million points)>

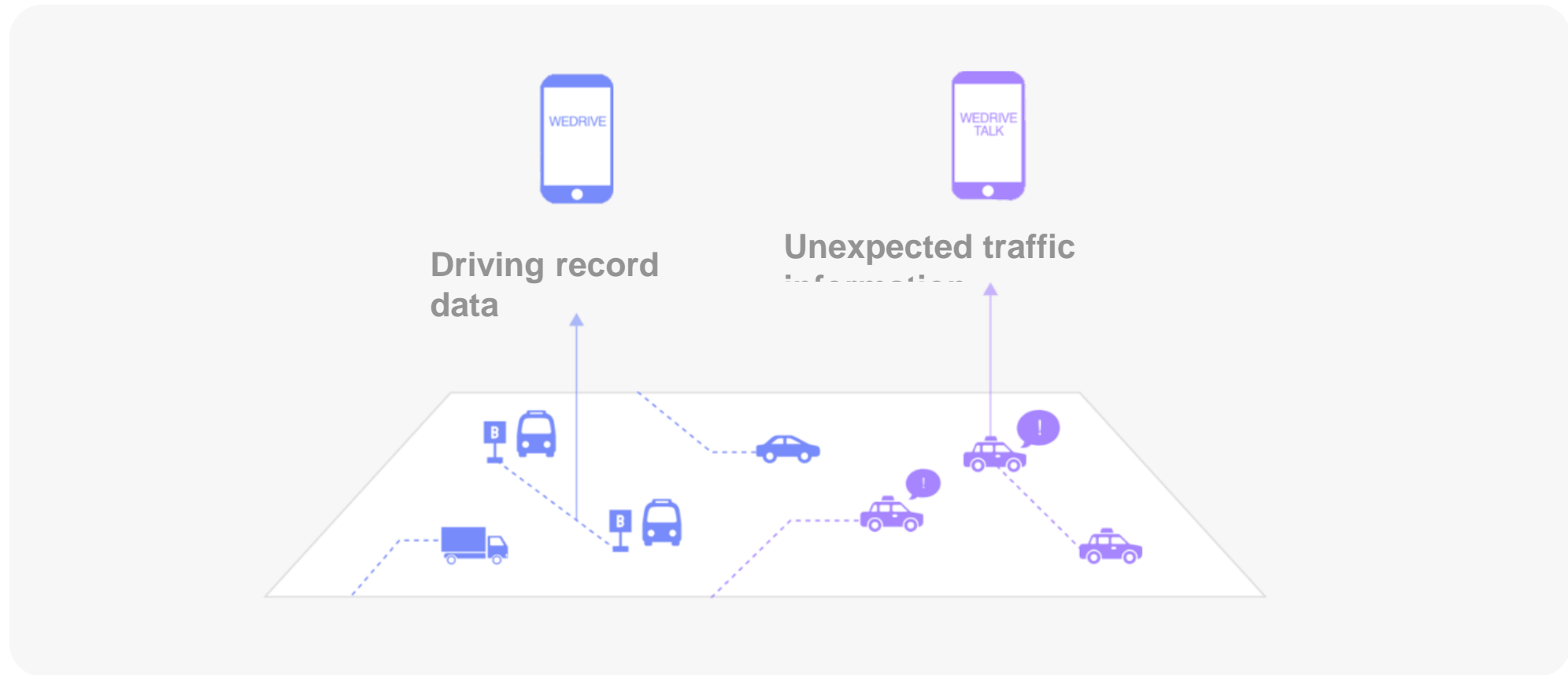
Introduction

| Objectives of this presentation

- Introduce a new VGI dataset collected by drivers in South Korea
- Present the role of time geography in detecting and addressing issues in the dataset
- Demonstrate an application of human mobility analysis using the new VGI dataset and potential applications

2. Data

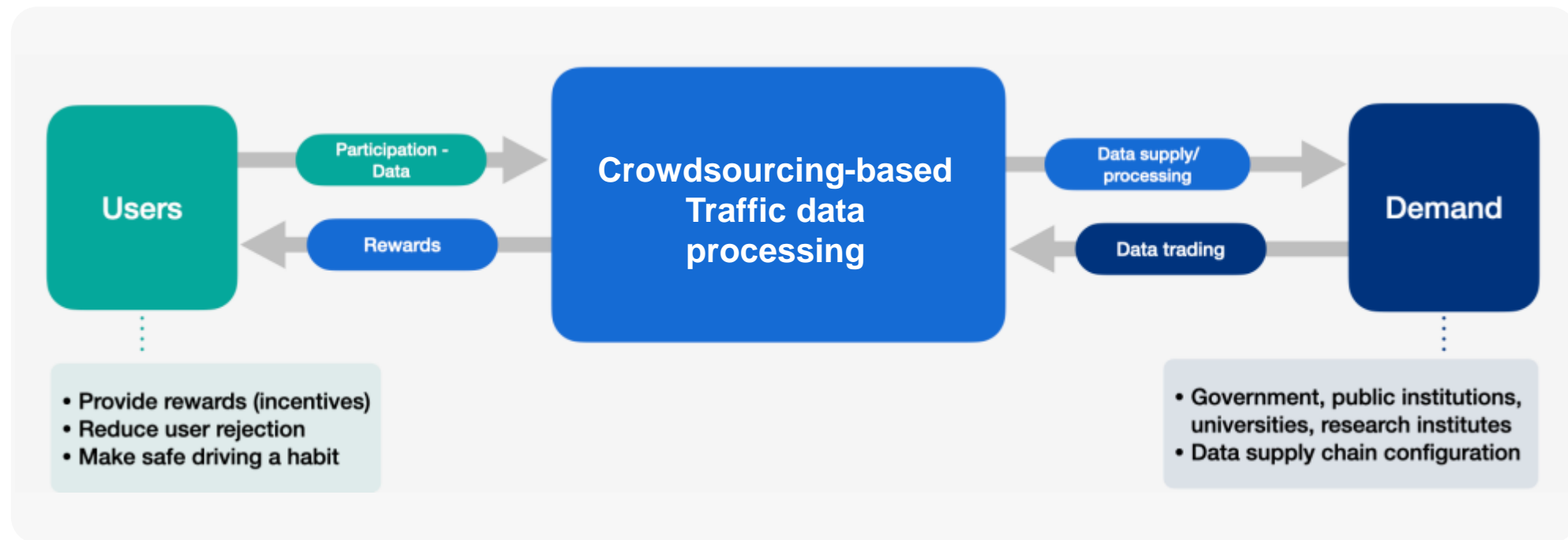
| VGI: Volunteered Geographic Information



Sources: WEDRIVE

<Citizen-driven transportation data collection>

VGI: Volunteered Geographic Information



Sources: WEDRIVE

<Eco-system of Data
management>

| Advantages of Smartphone based VGI datasets

01 Real-time traffic data collection

02 Longitudinal data that allows for relatively detailed time unit such as one month, two months, etc.

03 Enable to collect data where the areas with limited

04 Multimodal travel data analysis (i.e. walking, transit, personal vehicle)

VGI datasets

▼ Trip-level vehicle trajectory dataset

accuracy	id	lat	lng	speed	time	type
12.666	1257361	35.271816	128.8505374	9.353451	1587767320440	START
9.714	1257362	35.2720155	128.850258	15.819402	1587767322458	OVER_ACCEL
8.333	1257363	35.2723125	128.8499379	18.853914	1587767324686	NORMAL
6.8	1257364	35.2726124	128.8496614	21.114965	1587767326500	NORMAL
6.333	1257365	35.2729302	128.8493941	21.143787	1587767328461	NORMAL
6	1257366	35.2732372	128.8491357	20.504515	1587767330465	NORMAL
6	1257367	35.2735508	128.8488576	18.224524	1587767332530	NORMAL
6	1257368	35.2738109	128.8485909	19.486906	1587767334478	NORMAL
6	1257369	35.2740616	128.8482797	19.957775	1587767336469	NORMAL
6	1257370	35.274267	128.8479419	19.581053	1587767338470	NORMAL
6	1257371	35.274446	128.8475729	19.662083	1587767340463	NORMAL
6	1257372	35.2745888	128.8471918	19.228378	1587767342471	NORMAL
6	1257373	35.2746959	128.8468019	18.938776	1587767344459	NORMAL
5.6	1257374	35.2747661	128.8464104	18.078796	1587767346471	NORMAL
4.8	1257375	35.2748066	128.8460521	15.946122	1587767348491	NORMAL
4	1257376	35.2748362	128.8457737	13.006538	1587767350482	NORMAL

▼ Vehicle characteristics & mileage dataset

mid	carName	carNumber	distance
+QclpWW+E4E8WxMTJq++8K1ACGa8roGrD8OSME=	기아 뉴카렌스	21 33	226143439
+QclpWW+E4E8WxMTJq++Wlqg4sEzHY4Uj+8K0k7Y=	기아 봉고3	83 39	132634120
+QclpWW+E4E8WxMTJq++3q4qms2ZQqXSOQ014etY=	기아 K5	05 32	341912212
+QclpWW+E4E8WxMTJq++3k7pjsLq9RqT8QWb8rk=	기아 카니발	40 47	66689515
+QclpWW+E4E8WxMTJq++2J4U6d0Q0QOC31RLc5ys=	르노삼성 SM5	63 33	170001000
+QclpWW+E4E8WxMTJq++3grtK6EysTpsT8QpId8=	현대 쉐라퍼(승차용)	80 83	235367952
+QclpWW+E4E8WxMTJq++kZK51V1f8uqzC406v0=	세레스 이베로	61 16	95062745
+QclpWW+E4E8WxMTJq++8cTjxMJnGcWk2f5ccn=	현대 모터	92 74	139734931
+QclpWW+E4E8WxMTJq++8pV85HgBgczmR0k25SRQ=	현대 그랜지	11 53	15888592
+QclpWW+E4E8WxMTJq++c03pJqg3509y9huZDh4=	현대 그랜지	34 580	4817059
+QclpWW+E4E8WxMTJq++G3NGEncDF2meMNgZ2q2q=	현대 그랜지	60 54	352279965
+QclpWW+E4E8WxMTJq++gDRKXC8Zvrf089N8q=	기아 쏘렌토	15 26	17285000
+QclpWW+E4E8WxMTJq++GYNwNhy9pPLvdh8v=	86	16 31	42452443
+QclpWW+E4E8WxMTJq++hLp8Hy8E7L3NcW9Q29a=	기아 카니발	20 35	168772798
+QclpWW+E4E8WxMTJq++8wyfMR0K06s573p9w4=	현대 스타레츠	73 70	149643567
+QclpWW+E4E8WxMTJq++ioGLQ+ /6 +Y/MvH86BA8=	기아 리얼티	24 58	228869195
+QclpWW+E4E8WxMTJq++86 /qztmmM8R8gW8v8A=	2.5톤 다목적형	90 34	141322397
+QclpWW+E4E8WxMTJq++J2Hxlp075Z6509fA8RM=	현대 i40	84 14526	237527264
+QclpWW+E4E8WxMTJq++8 /m9pL0Q3H5uab0K7pY=	르노	56 73	259751303
+QclpWW+E4E8WxMTJq++JReQQTdZfhrNCTVQqA=	BMW 5시리즈	49 33	99162183
+QclpWW+E4E8WxMTJq++LLY/LJXMI7N7pGCoZ2dW4H=	BMW GT3	63 54	104280612
+QclpWW+E4E8WxMTJq++62D00N33ka7V9344aE=	기아 그랜지 카니발	42 35	16953656
+QclpWW+E4E8WxMTJq++M1ST13LD8eHfZc593ys=	르노삼성 QM6	16 171	653133
+QclpWW+E4E8WxMTJq++NaYVCx5aYD8RA98TTE=	BMW 1시리즈	40 83	82478613
+QclpWW+E4E8WxMTJq++85gqTpsKv9pTK8gme=	현대 코나	89 26	18704080
+QclpWW+E4E8WxMTJq++01UyK38JQhQoc5aMATE=	현대 엑센트	27 36	43042151
+QclpWW+E4E8WxMTJq++08t+AV2GMN85tVh8AU=	PRIMA	22 12355	225627318
+QclpWW+E4E8WxMTJq++vV49JwC3AeDQ193AU=	기아 K5	21 67	178590005
+QclpWW+E4E8WxMTJq++8Vez8MyCu3Z8K5D51VY=	기아 스포티지	65 52	143291496
+QclpWW+E4E8WxMTJq++rggdiHcKCu3StnDTG2H=	기아 K7	20 75	60050074
+QclpWW+E4E8WxMTJq++K48FHG2YC4gUJURIGcl=	쌍미 쉐넬	40 36	77107994
+QclpWW+E4E8WxMTJq++RYQpDeHwqy57b89H4Z4=	기아 스포티지	51 32	105409945
+QclpWW+E4E8WxMTJq++dANLdpKK8AV8y8pK8Yc=	현대 스타레츠	70 87	248351274

▼ Trip-level origin-destination (O-D) dataset

time_begin	time_end	distance	rank	score	heart	rapid0	rapid1	night	origin_lat	origin_lng	destination_lat	destination_lng
2020-06-18 07:00:59	2020-06-18 07:04:51	2750	5	0	2	2	0	0	35.3399	128.796	35.3141	128.728
2020-06-18 07:05:02	2020-06-18 09:09:39	55512	5	0	55	55	0	0	35.3122	128.728	35.3058	129.027
2020-06-18 09:53:43	2020-06-18 09:59:31	3251	5	0	3	3	0	0	35.313	129.628	35.3404	129.034
2020-06-18 09:59:37	2020-06-18 10:07:32	4168	5	0	4	4	0	0	35.341	129.034	35.3692	129.052
2020-06-18 10:07:37	2020-06-18 10:18:18	1133	5	0	1	1	0	0	35.369	129.053	35.3714	129.056
2020-06-18 10:07:37	2020-06-18 11:09:36	5261	5	0	5	5	0	0	35.368	129.054	35.3488	129.046
2020-06-18 11:09:38	2020-06-18 11:19:34	11296	5	0	11	11	0	0	35.3488	129.046	35.2709	128.982
2020-06-18 11:19:50	2020-06-18 12:19:58	28774	5	0	28	28	0	0	35.2702	128.976	35.3301	128.708
2020-06-18 12:42:15	2020-06-18 12:48:02	1700	5	0	1	1	0	0	35.3308	128.709	35.3442	128.71
2020-06-18 12:57:46	2020-06-18 13:22:00	12752	5	0	12	12	0	0	35.3441	128.71	35.2873	128.791
2020-06-18 13:30:02	2020-06-18 13:31:18	1407	5	0	1	1	0	0	35.2262	128.835	35.2148	128.84
2020-06-18 13:31:47	2020-06-18 13:34:36	3181	5	0	3	3	0	0	35.2132	128.846	35.2121	128.88
2020-06-18 13:35:42	2020-06-18 13:38:23	2584	5	0	2	2	0	0	35.2142	128.894	35.2174	128.918
2020-06-18 13:38:34	2020-06-18 14:08:07	5186	5	0	5	5	0	0	35.2183	128.92	35.2179	128.918
2020-06-18 14:06:39	2020-06-18 14:08:55	2361	5	0	2	2	0	0	35.2162	128.912	35.2132	128.887
2020-06-18 14:09:06	2020-06-18 14:17:17	5728	5	0	5	5	0	0	35.2128	128.884	35.2065	128.831
2020-06-18 14:17:33	2020-06-18 14:23:05	3747	5	0	3	3	0	0	35.2059	128.829	35.1903	128.793
2020-06-18 14:23:25	2020-06-18 14:28:08	2625	5	0	2	2	0	0	35.1882	128.761	35.1823	128.767
2020-06-18 14:26:46	2020-06-18 14:31:21	5585	5	0	5	5	0	0	35.1818	128.761	35.1986	128.704
2020-06-18 14:33:37	2020-06-18 16:01:44	30110	5	0	30	30	0	0	35.2008	128.7	35.2008	128.918
2020-06-18 16:18:33	2020-06-18 16:21:00	1079	5	0	1	1	0	0	35.2291	128.92	35.2224	128.927
2020-06-18 16:22:07	2020-06-18 16:29:35	8046	5	0	8	8	0	0	35.2189	128.922	35.2172	128.838
2020-06-18 16:31:28	2020-06-18 16:34:35	3493	5	0	3	3	0	0	35.2354	128.833	35.2398	128.811
2020-06-18 16:37:06	2020-06-18 16:48:29	1549	5	0	1	1	0	0	35.2821	128.8	35.2853	128.79
2020-06-18 16:42:29	2020-06-18 16:48:18	2111	5	0	2	2	0	0	35.2872	128.788	35.2939	128.797
2020-06-18 17:37:50	2020-06-18 17:49:55	2235	5	0	2	2	0	0	35.291	128.812	35.2925	128.793
2020-06-18 17:41:06	2020-06-18 17:41:45	1494	5	0	1	1	0	0	35.2921	128.791	35.2952	128.775
2020-06-18 17:43:50	2020-06-18 17:58:48	5222	5	0	5	5	0	0	35.2955	128.775	35.3075	128.724
2020-06-18 17:54:53	2020-06-18 18:07:00	3612	5	0	3	3	0	0	35.308	128.724	35.2999	128.708
2020-06-18 18:10:31	2020-06-18 18:30:55	12862	A	0	12	12	0	1	37.2869	126.853	37.2125	126.817
2020-06-18 18:28:36	2020-06-18 18:24:24	3028	A	0	3	3	0	1	37.2187	126.825	37.2015	126.829
2020-06-18 18:40:36	2020-06-18 18:45:42	1649	5	0	1	1	0	0	37.21	126.826	37.2125	126.817
2020-06-18 18:42:10	2020-06-18 18:43:43	11789	C	0	11	5	0	7	37.2197	126.827	37.2067	126.853
2020-06-18 08:50:28	2020-06-18 07:07:25	3658	5	0	5	5	0	0	34.8898	128.875	34.8884	128.706
2020-06-18 08:27:03	2020-06-18 08:24:32	2955	5	0	2	2	0	0	34.8749	128.707	34.8797	128.717
2020-06-18 08:33:08	2020-06-18 08:42:47	1902	5	0	1	1	0	0	34.8787	128.717	34.8799	128.717

Sources: WEDRIVE

| Data overview

**Number of accumulated users / MAU /
DAU**

**Total cumulatively collected travel distance /
Number of trips / Volume of dataset**

**Daily average of collected travel distance /
Number of trips / Volume of dataset**



200K / 40K / 24K

184M(km) / 83.25M / 8.5TB

2M(km) / 100K / 15GB

Sources: WEDRIVE

Data

Pre-processing : Role of Time Geography in detecting and addressing issues in VGI datasets

Developed a data preprocessing algorithm based on Space-Time Prism for 546 days

Table of (a)

Dep. Time	Dep. Lon.	Dep. Lat.	Arr. Time	Arr. Lon.	Arr. Lat.
10:00	128°E	35.8°N	12:00	129°E	36°N
15:00	129°E	36°N	16:00	127.55°E	36°N
18:00	127.55°E	36°N	26:00	128°E	35.8°N

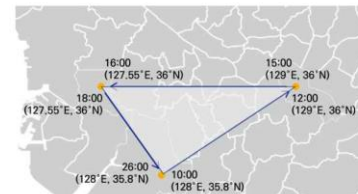
Sequence process from (a)

Time	Lon.	Lat.
00:00	129°E	36°N
9:59	129°E	36°N
10:00	128°E	35.8°N
12:00	129°E	36°N
...
26:00	128°E	35.8°N
25:59
26:01	128°E	35.8°N
...
27:59	128°E	35.8°N

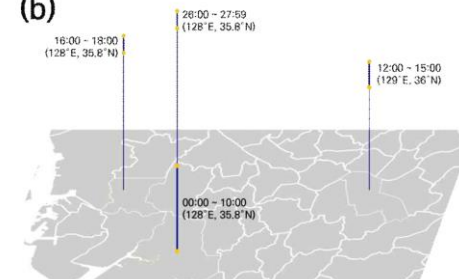
Table of (c)

Time	Lon.	Lat.
00:00	128°E	35.8°N
00:01	128°E	35.8°N
00:02	128°E	35.8°N
...
10:00	128°E	35.8°N
10:01	128.00833°E	35.80167°N
10:02	128.01667°E	35.80333°N
10:03	128.025°E	35.805°N
...
27:57	128°E	35.8°N
27:58	128°E	35.8°N
27:59	128°E	35.8°N

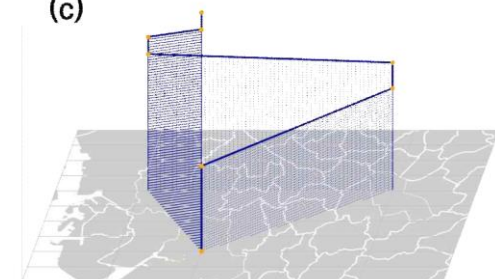
(a)



(b)



(c)



Pre-processing : Role of Time Geography in detecting and addressing issues in VGI datasets

Date	Dep. Time	Dep. Lon.	Dep. LAT.	Arr. Time	Arr. Lon.	Arr. LAT.
2021. 1. 1.	10:00	128.40°E	35.60°N	11:00	128.70°E	35.80°N
2021. 1. 1.	15:00	128.70°E	35.80°N	16:00	128.35°E	36°N
2021. 1. 1.	22:00	128.35°E	36°N	23:00	128.40°E	35.60°N
...						
2021. 9. 18.	12:00	128.°E	35.90°N	13:00	128.75°E	35.70°N
2021. 9. 18.	16:00	128.75°E	35.70°N	17:00	128.60°E	36°N
2021. 9. 18.	18:00	128.60°E	36°N	19:00	128.55°E	35.85°N
2021. 9. 18.	22:00	128.55°E	35.85°N	23:00	128°E	35.90°N
...						
2022. 6. 30.	08:00	128.35°E	36°N	09:00	128.60°E	35.80°N
2022. 6. 30.	12:00	128.60°E	35.90°N	13:00	128.50°E	35.75°N
2022. 6. 30.	15:00	128.50°E	35.75°N	16:00	128.35°E	36°N

525 days, 6,718 trips



Date	Time	Lon.	Lat.
2021. 1. 1.	00:00	128.40°E	35.60°N
2021. 1. 1.	00:01	128.40°E	35.60°N
...			
2021. 1. 1.	10:00	128.40°E	35.60°N
2021. 1. 1.	10:01	128.40508°E	35.60339°N
2021. 1. 1.	10:02	128.41017°E	35.60678°N
...			
2021. 1. 1.	23:58	128.40°E	35.60°N
2021. 1. 1.	23:59	128.40°E	35.60°N
...			
2022. 6. 30.	00:00	128.35°E	36°N
2022. 6. 30.	00:01	128.35°E	36°N
...			
2022. 6. 30.	12:00	128.60°E	35.90°N
2022. 6. 30.	12:01	128.59830°E	35.89746°N
2022. 6. 30.	12:02	128.59661°E	35.89492°N
...			
2022. 6. 30.	23:58	128.35°E	36°N
2022. 6. 30.	23:59	128.35°E	36°N

546(day) * 24(hour) * 60(min) = 786,240

| Five issues in smartphone based VGI datasets

(1) Incomplete records of an individual's movement(activity locations)

Before the trip
(00:00~15:57)

Case.

UUID	Departure(kst)	Arrival(kst)	Distance(m)	Orgin_lat(°)	Origin_Ing(°)	Destination_lat(°)	Destination_Ing(°)
003bc278fz	2021-01-01 15:58:06	2021-01-01 16:30:18	8,675	35.846	128.538	35.850	128.528
003bc278fz	2021-01-01 16:39:38	2021-01-01 17:38:41	31,802	35.850	128.528	36.035	128.623
003bc278fz	2021-01-01 20:33:45	2021-01-01 21:27:26	36,503	36.035	128.623	35.846	128.538

| Five issues in smartphone based VGI datasets

(1) Incomplete records of an individual's movement(activity locations)

Case.

UUID	Departure(kst)	Arrival(kst)	Distance(m)	Orgin_lat(°)	Origin_Ing(°)	Destination_lat(°)	Destination_Ing(°)
003bc278fz	2021-01-01 15:58:06	2021-01-01 16:30:18	8,675	35.846	128.538	35.850	128.528
003bc278fz	2021-01-01 16:39:38	2021-01-01 17:38:41	31,802	35.850	128.528	36.035	128.623
003bc278fz	2021-01-01 20:33:45	2021-01-01 21:27:26	36,503	36.035	128.623	35.846	128.538

After the trip (21:27~23:59)

Five issues in smartphone based VGI datasets

(1) Incomplete records of an individual's movement(activity locations)

Case.

UUID	Departure(kst)	Arrival(kst)	Distance(m)	Orgin_lat(°)	Origin_Ing(°)	Destination_lat(°)	Destination_Ing(°)
003bc278fz	2021-01-01 15:58:06	2021-01-01 16:30:18	8,675	35.846	128.538	35.850	128.528
003bc278fz	2021-01-01 16:39:38	2021-01-01 17:38:41	31,802	35.859	128.527	36.035	128.623
003bc278fz	2021-01-01 20:33:45	2021-01-01 21:27:26	36,503	36.035	128.623	35.846	128.538

**Between the trips
(16:30~16:39 & 17:38~20:33)**

| Five issues in smartphone based VGI datasets

(2) Same departure and arrival time

Case 1.

UUID	Departure(kst)	Arrival(kst)	Distance(m)
200df23393	2022-04-20 12:34:40	2022-04-20 13:53:21	2,221
200df23393	2022-04-20 12:34:40	2022-04-20 17:13:50	5,032

Case 2.

UUID	Departure(kst)	Arrival(kst)	Distance(m)
110234aadf	2021-01-03 12:34:40	2022-01-03 18:33:01	4,032
110234aadf	2021-01-03 15:33:21	2022-01-03 18:33:01	2,303

| Five issues in smartphone based VGI datasets

(3) Nested trip within the preceding trip

Case.

UUID	Departure(kst)	Arrival(kst)	Distance(m)
00eq78aadf	2021-01-03 12:45:40	2022-01-03 18:33:01	4,769
00eq78aadf	2021-01-03 16:17:58	2022-01-03 17:53:01	1,501

| Five issues in smartphone based VGI datasets

(4) Extremely low or high speed

Case.

UUID	Departure(kst)	Arrival(kst)	Distance(m)	Duration
96e2e09b0	2021-01-01 20:18:12	2021-01-02 09:10:53	1,458	46,325s (~12.87 hours)
84efd934ac	2021-01-02 10:01:21	2021-01-06 19:41:01	34,479	380,421s (~4.4 days)
906941c595	2021-01-02 12:29:04	2021-01-03 12:01:33	1,991	84,691s (~23.53 hours)
e6a2b04d5c	2021-01-02 18:13:54	2021-01-03 12:04:23	1,112	64,279s (~17.86 hours)

| Five issues in smartphone based VGI datasets

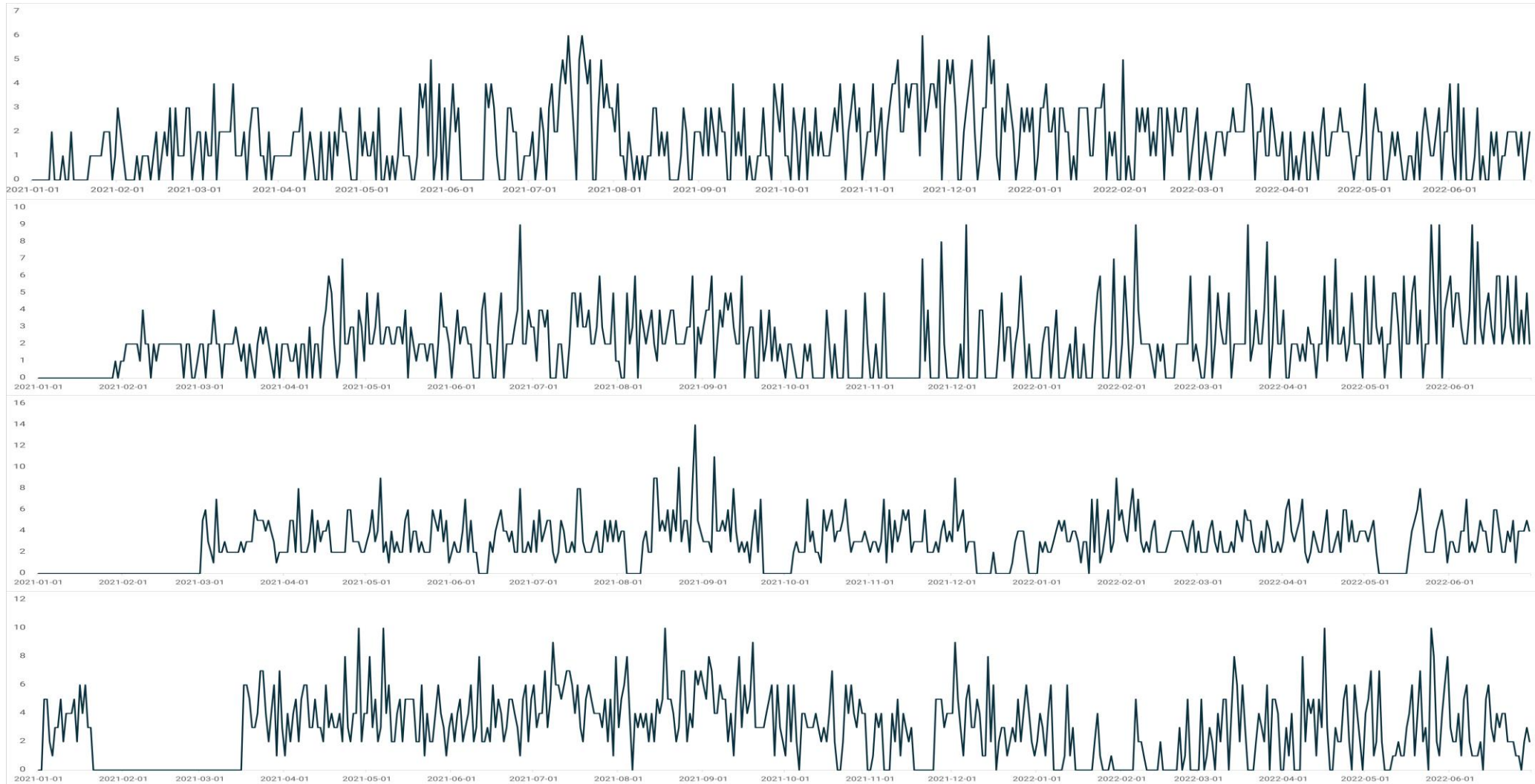
(5) Lack of regularity

Case.

UUID	Departure(kst)	Arrival(kst)
781bdk586z	2021-01-01 15:13:06	2021-01- 01 15:26:03
781bdk586z	2021-01-06 18:52:39	2021-01- 06 18:59:16
781bdk586z	2021-01-09 19:25:18	2021-01- 09 20:14:02
781bdk586z	2021-01-19 15:19:57	2021-01- 19 15:26:50
781bdk586z	2021-01-21 17:23:40	2021-01- 21 17:43:03
781bdk586z	2021-01-22 00:37:29	2021-01- 22 00:45:57

Only 6 days (in 546 days)

Irregularity : Daily Travel Frequency Distribution of 4 respondents



Data

Pre-processing

So, we implemented a preprocessing algorithm to impute incomplete records(1), to filter our error cases(2-4).

Table of (a)

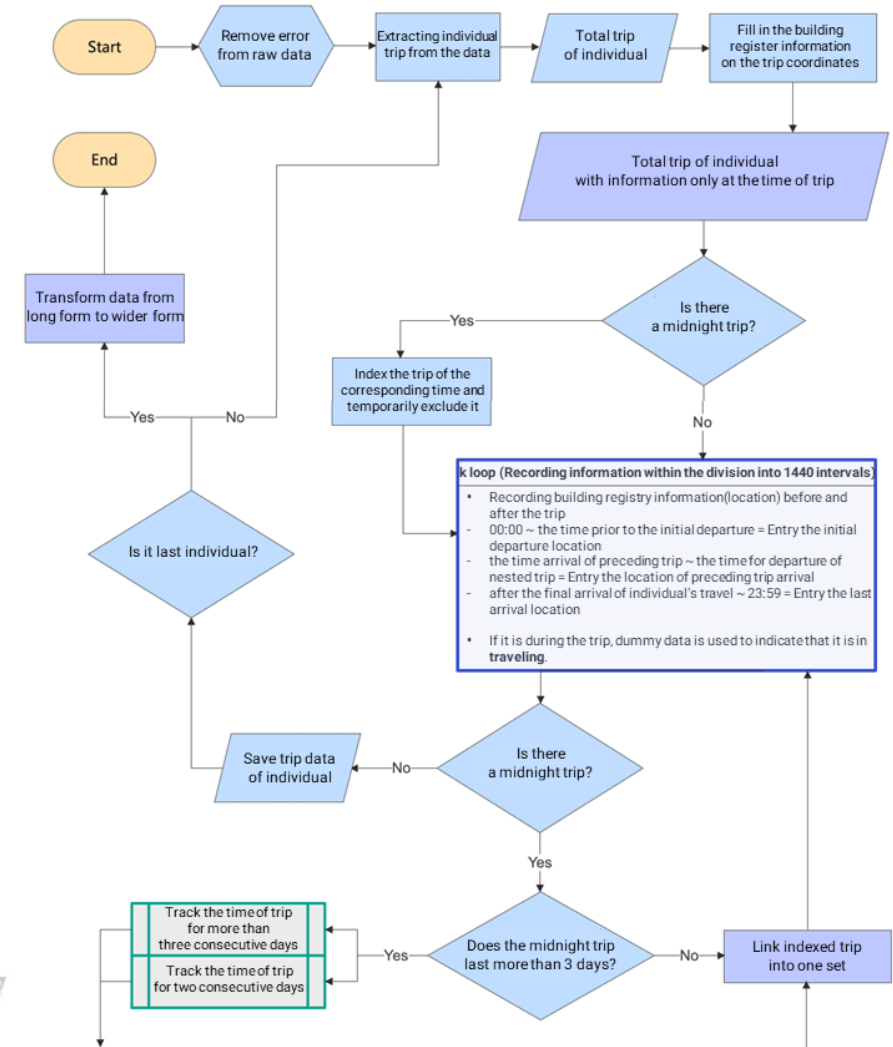
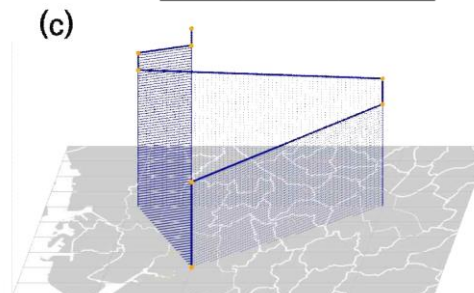
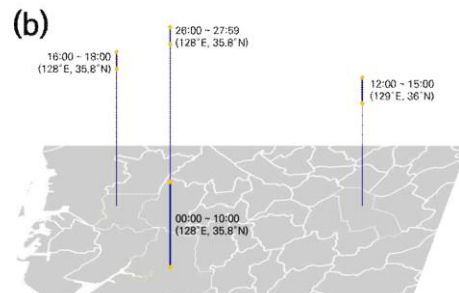
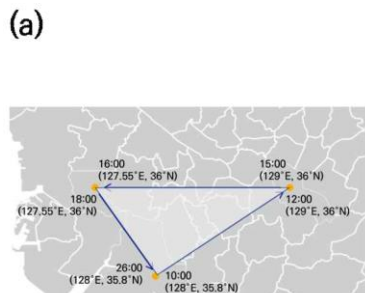
Dep. Time	Dep. Lon.	Dep. Lat.	Arr. Time	Arr. Lon.	Arr. Lat.
10:00	128°E	35.8°N	12:00	129°E	36°N
15:00	129°E	36°N	16:00	127.55°E	36°N
18:00	127.55°E	36°N	26:00	128°E	35.8°N

Sequence process from (a)

Time	Lon.	Lat.
00:00	129°E	36°N
...
9:59	129°E	36°N
12:01		Trip
...
14:59		Trip
...
25:59		Trip
26:01	128°E	35.8°N
...
27:59	128°E	35.8°N

Table of (c)

Time	Lon.	Lat.
00:00	128°E	35.8°N
00:01	128°E	35.8°N
00:02	128°E	35.8°N
...
10:00	128°E	35.8°N
10:01	128.00833°E	35.80167°N
10:02	128.01667°E	35.80333°N
10:03	128.025°E	35.805°N
...
27:57	128°E	35.8°N
27:58	128°E	35.8°N
27:59	128°E	35.8°N

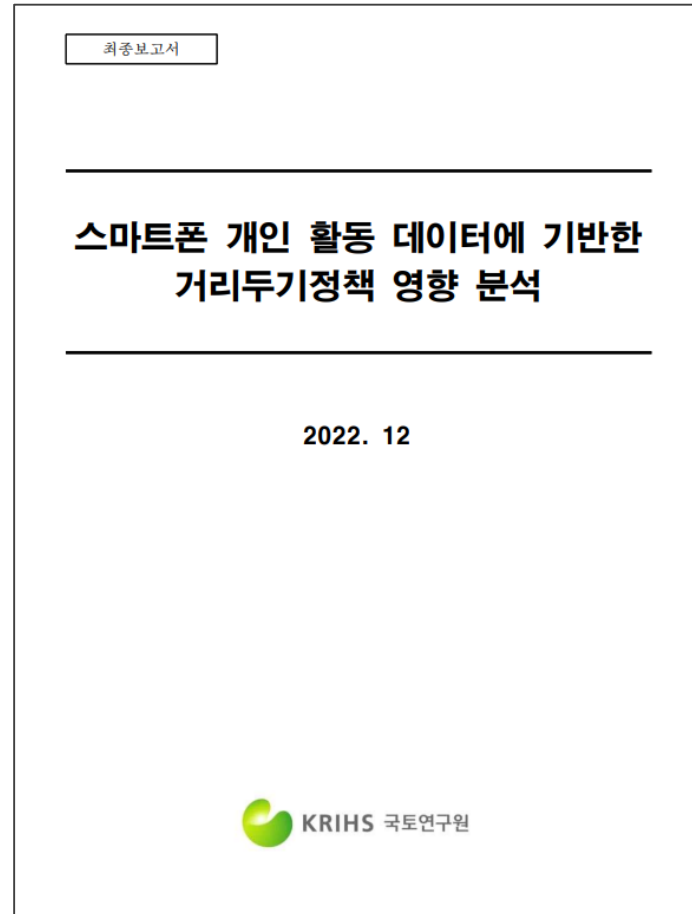


<Algorithm Flowchart>

3. Case study & Potential Applications

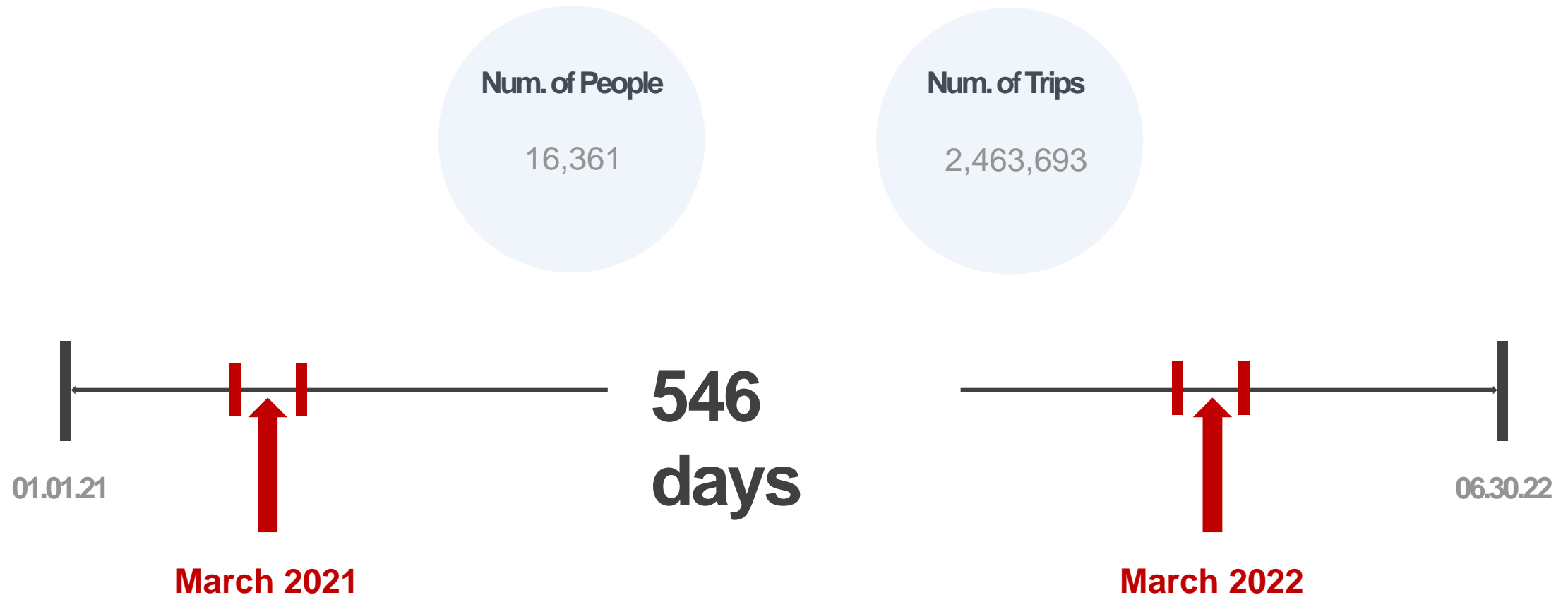
Case study

| Analyzing the effect of social-distancing policies based on the smartphone based VGI dataset



Case study

Data

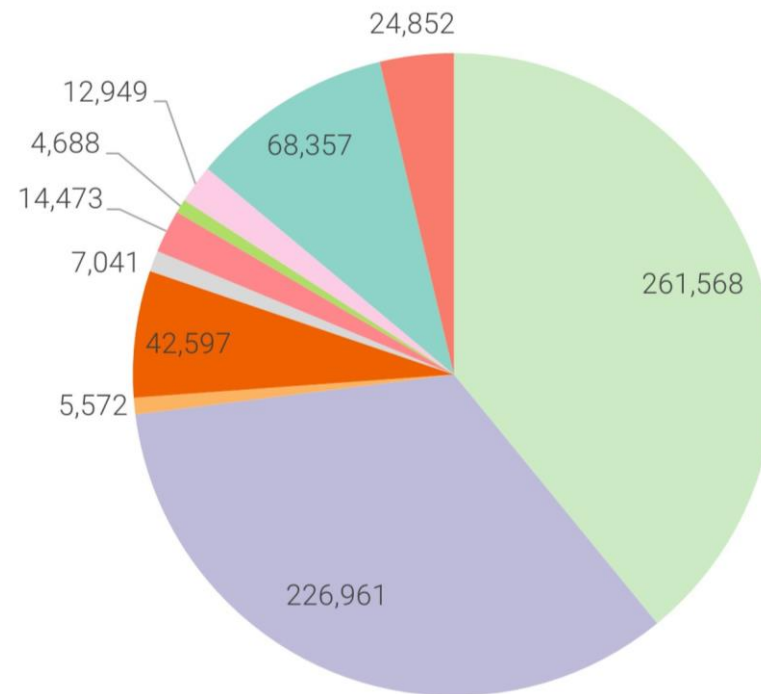


In this case study, **98,609** of trips collected by **479** individuals were used
The 479 individuals are the people provided their location information for more than
370 days (total number of weekdays in 546 days)

Case study

Methodology

(1) Land use characteristics – Spatial join(Million by Million)



Sources: Yoon et al. (2022)

- Activity Location
- 50-meter radius
- Residential areas
- Neighborhood living facilities
- Cultural & Sports
- Commercial
- Medical
- Education & Research
- Hotels
- Business
- Factory & Garage
- Other
- Out of Daegu
- Traveling

Case study

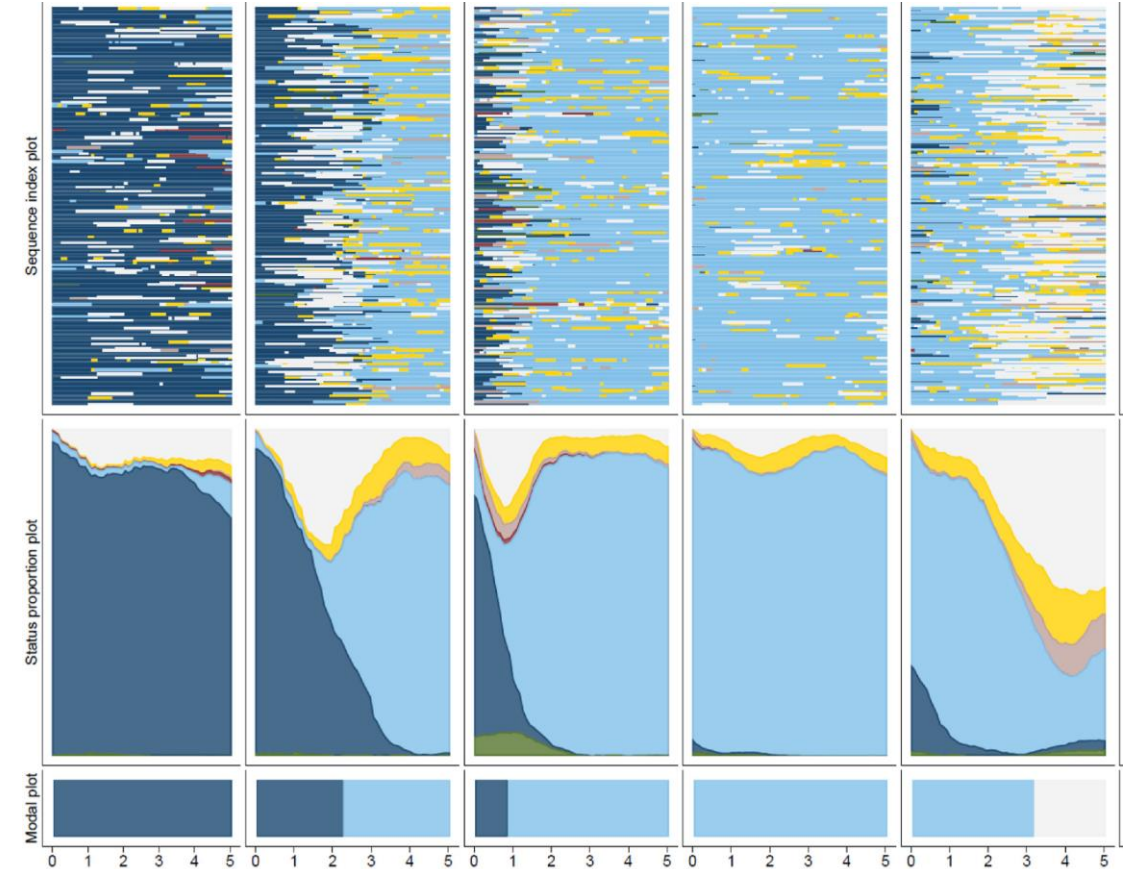
Methodology

(2) Sequence analysis

$$p(s_j \mid s_i) = \frac{\sum_{t=1}^{L-1} n_{t,t+1}(s_i, s_j)}{\sum_{t=1}^{L-1} n_t(s_i)}$$

$$SC(S_p, S_q) = 2 - P(S_p|S_q) - P(S_q|S_p)$$

$$D_Matrix = \begin{bmatrix} 0 & d_{x_1, x_2} & \cdots & d_{x_1, x_N} \\ d_{x_2, x_1} & 0 & \cdots & d_{x_2, x_N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{x_N, x_1} & d_{x_N, x_2} & \cdots & 0 \end{bmatrix} (d_{x_i, x_j} = d_{x_j, x_i} (i, j \in [1, N]))$$



Sources: Brzinsky-Fay et al. (2016)

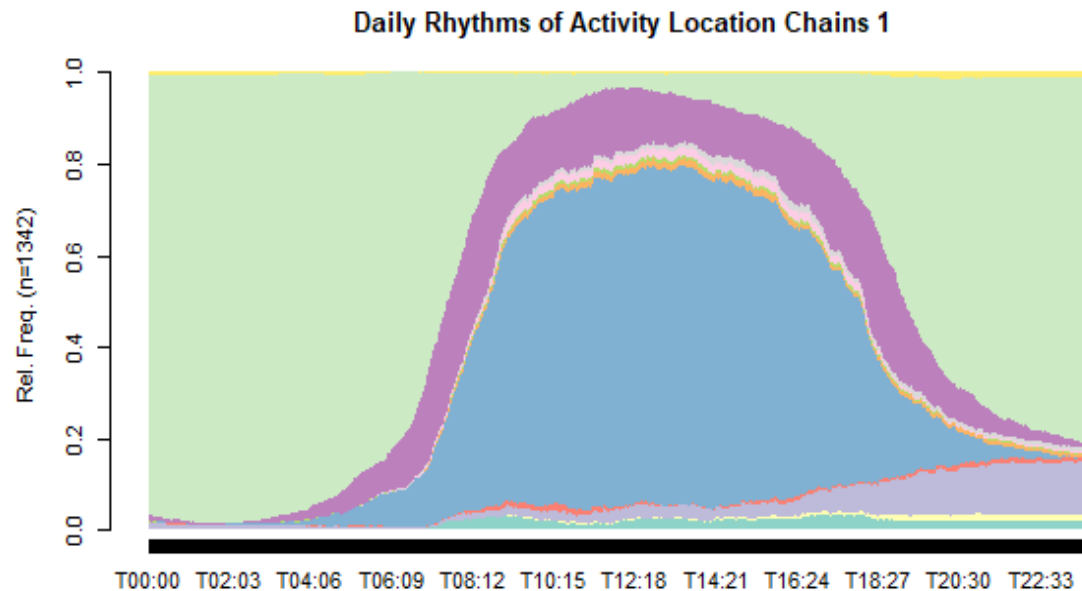
Case study

Results : Activity location logs

(2) Sequence analysis based on residents of Daegu Metropolitan City

Sources: Yoon et al. (2022)

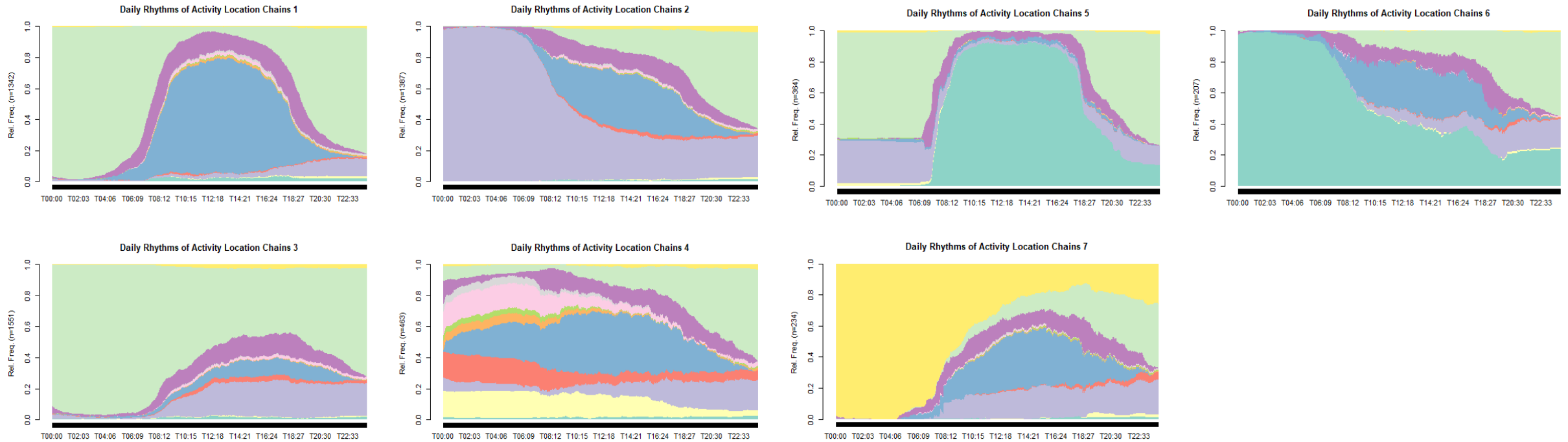
Example)



- Residential areas
- Neighborhood living facilities
- Cultural & Sports
- Commercial
- Medical
- Education & Research
- Hotels
- Business
- Factory & Garage
- Other
- Out of Daegu
- Traveling

Case study

Results : Activity location logs

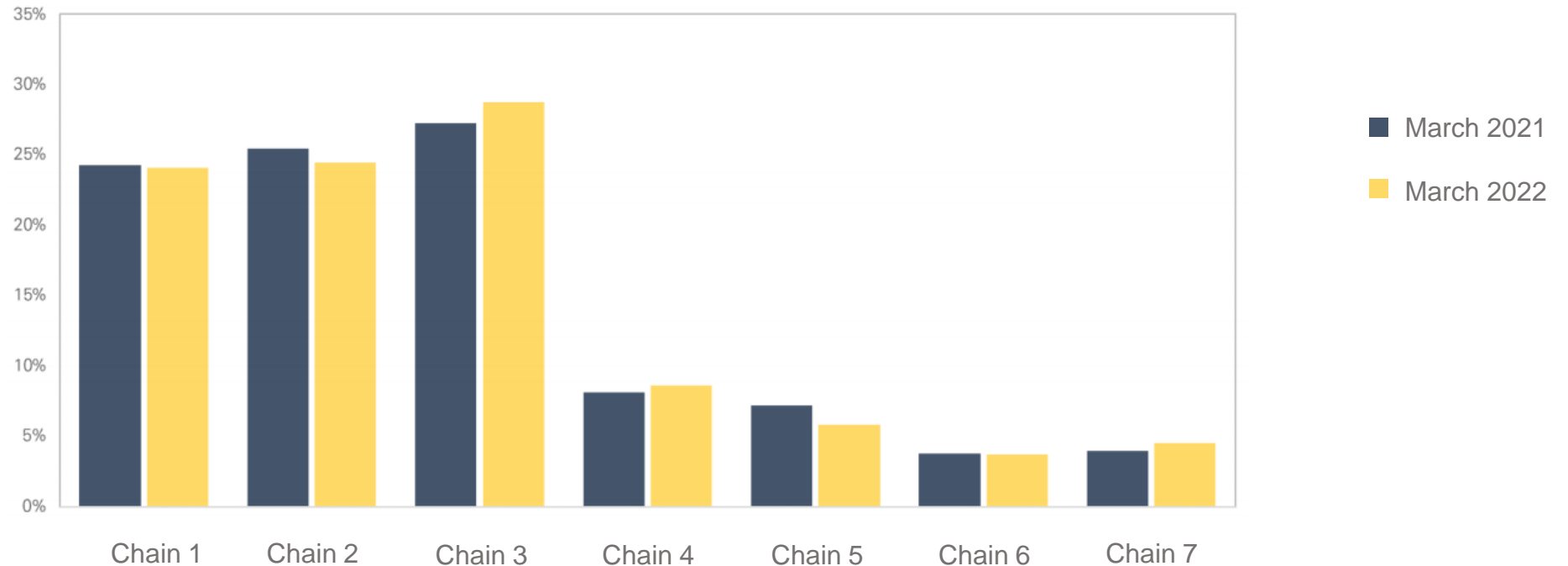


Case study

Results

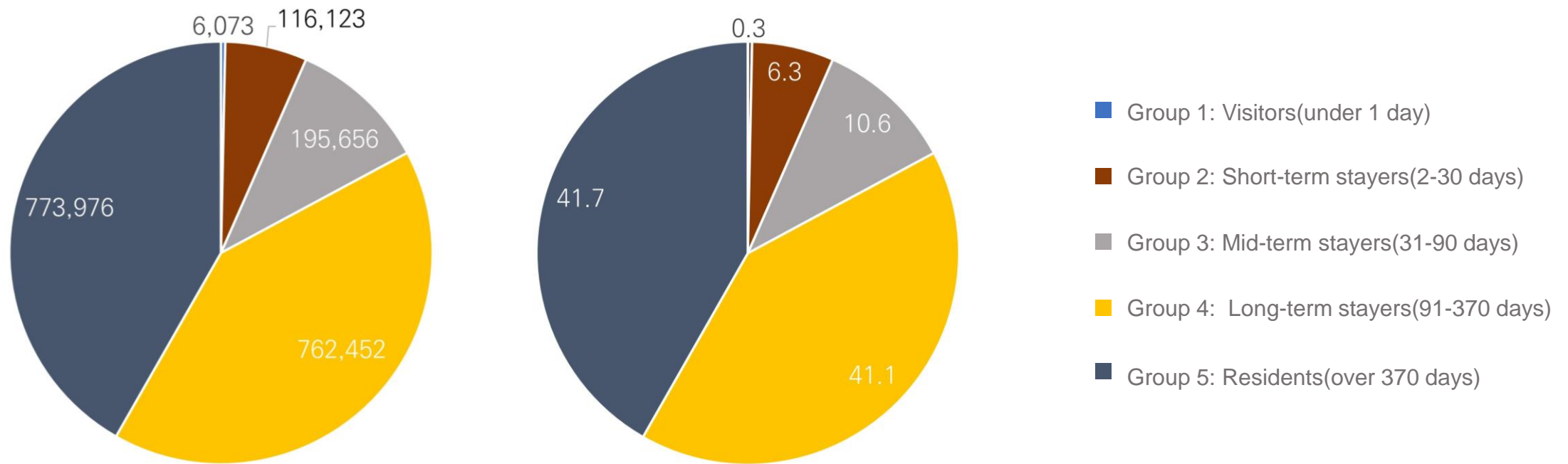
(3) Comparing the proportion of activity-location chains in March 2021(during the peak spread of COVID-19) to March 2022

Sources: Yoon et al. (2022)



Potential Applications 1

Visitors vs Stayers vs Residents

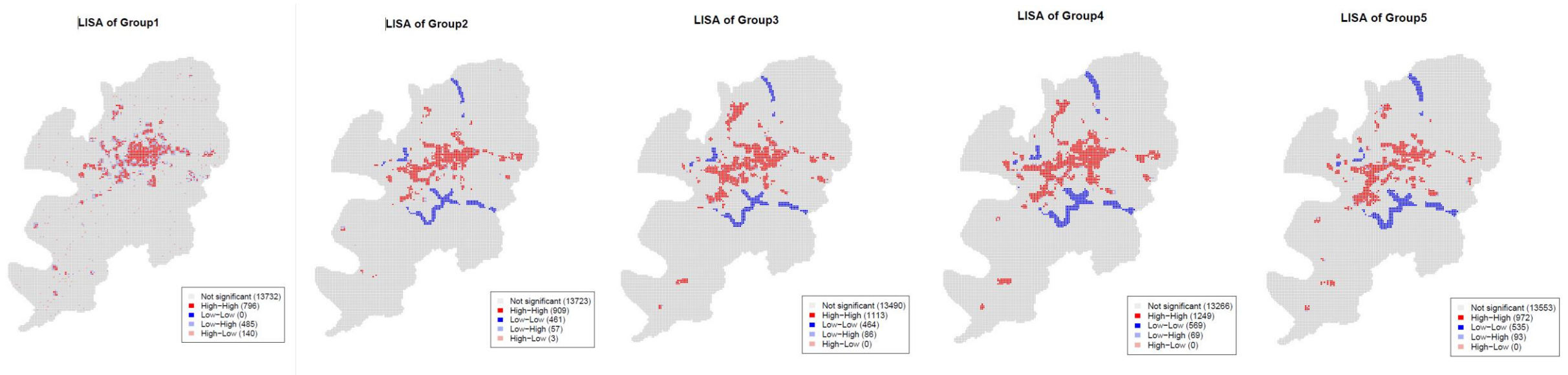


Sources: Yoon et al. (2022)

<Group classification based on weekdays of observation (left: N, right : Proportion)>

Potential Applications 1

Visitors vs Stayers vs Residents : LISA(Local Indicators of Spatial Association)

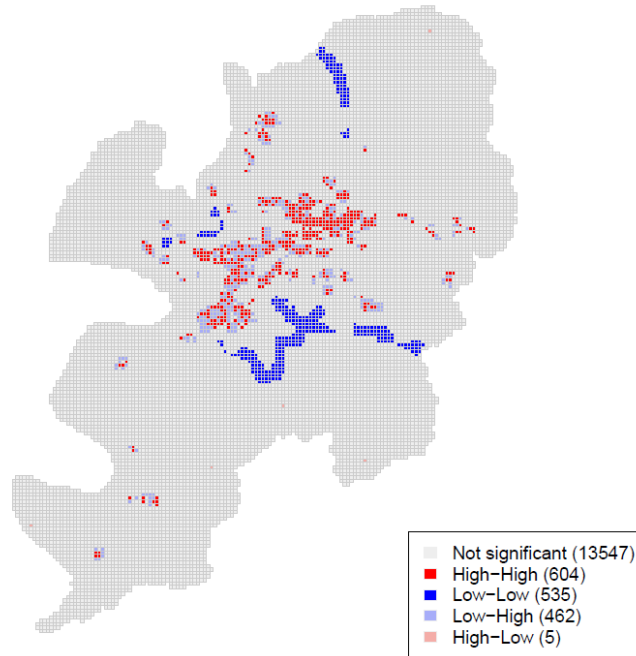


<Group-specific univariate LISA analysis>

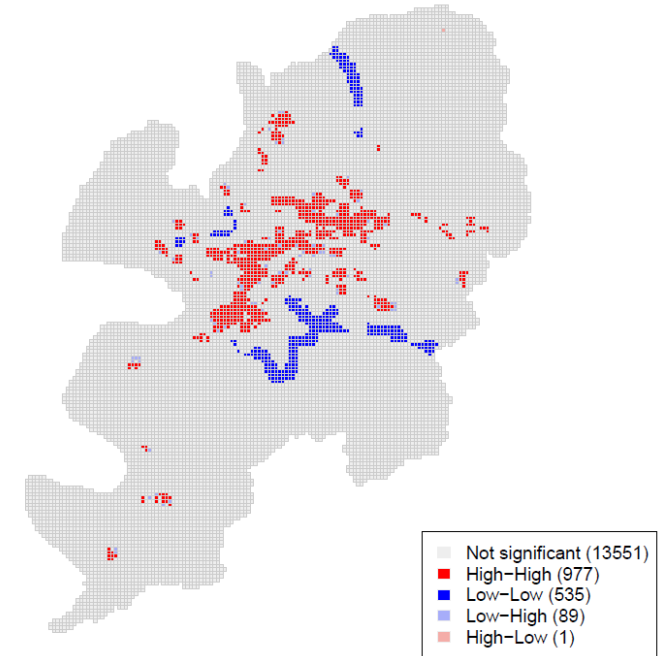
Potential Applications 1

Visitors vs Stayers vs Residents : LISA(Local Indicators of Spatial Association)

LISA of Group1 with Group5



LISA of Group2 with Group5

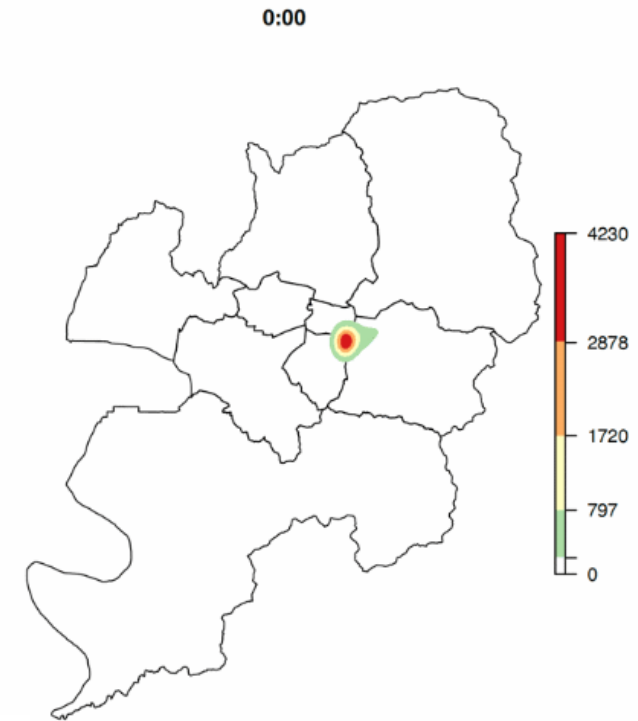
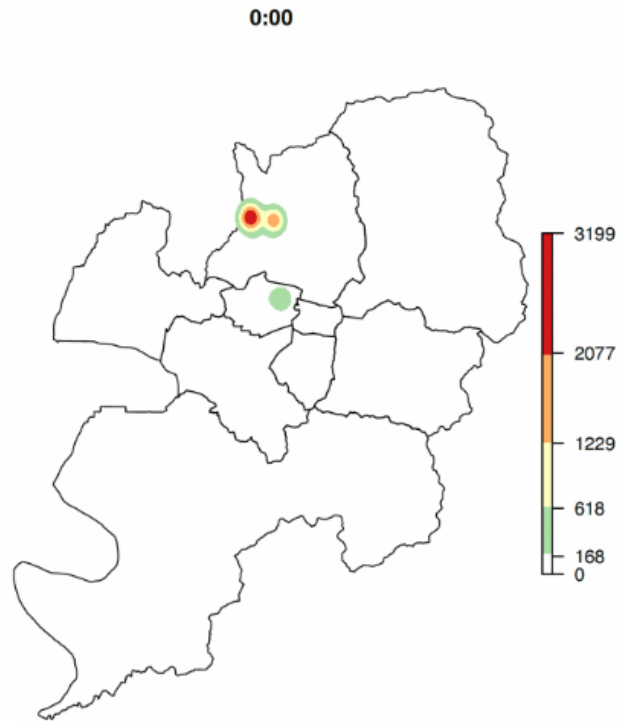


<Inter-group bivariate LISA analysis>

Potential Applications 2

| Density based activity space estimation

(1) Kernel Density Estimation

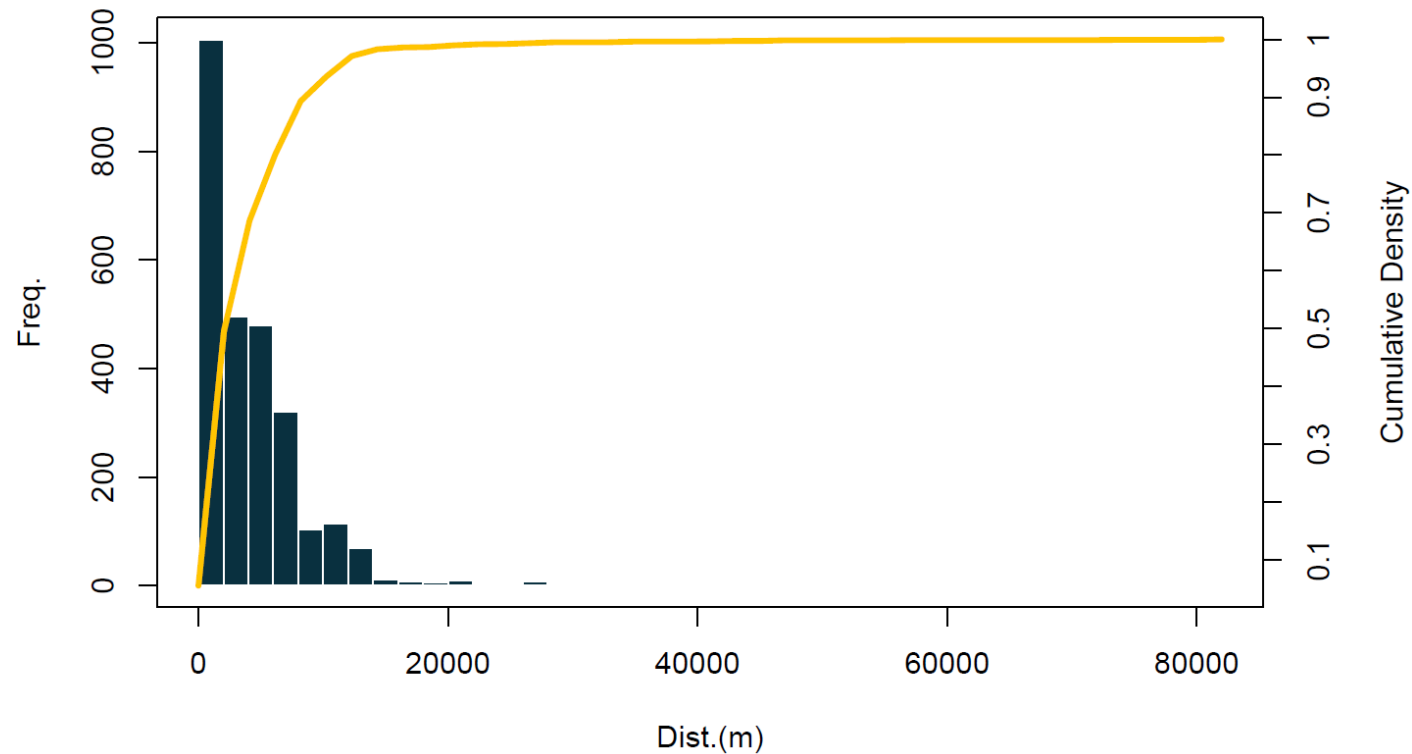


<Kernel Density Estimation by time period for destination coordinates>

Potential Applications 3

Examining the first law of geography in a person level

(1) ECDF(Empirical Cumulative Distribution Function)

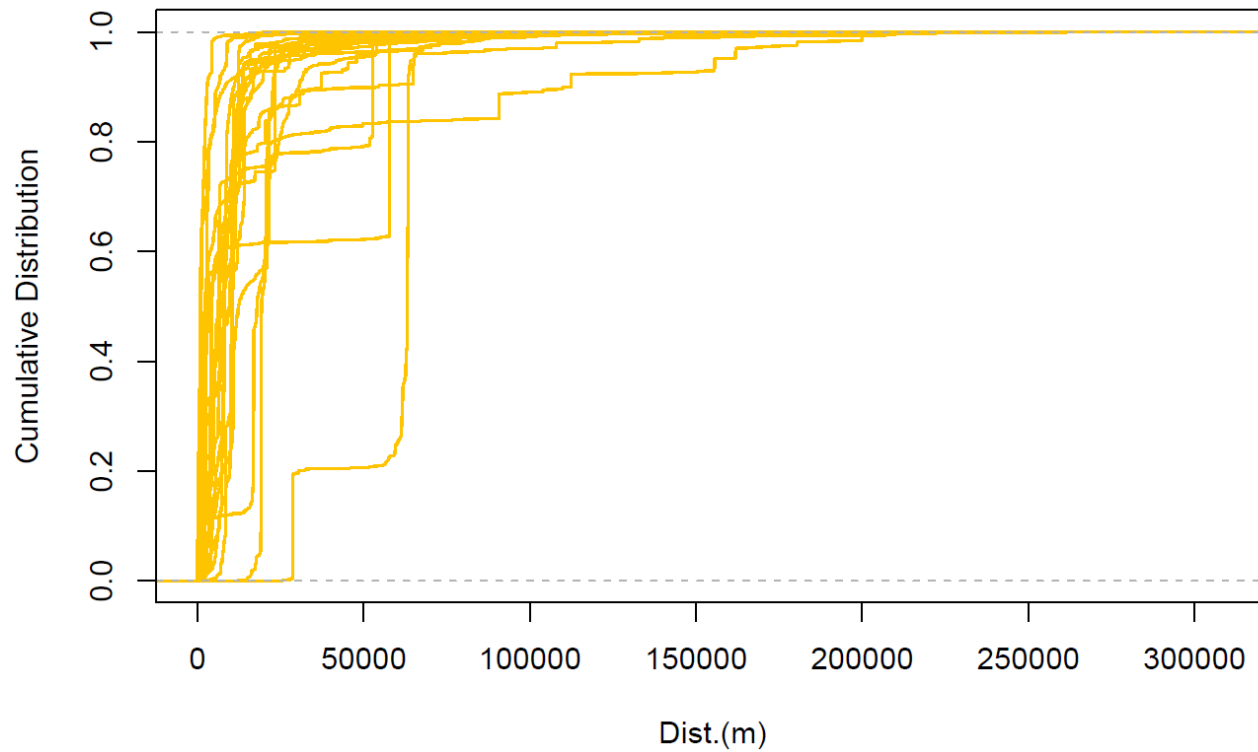


<ECDF of travel distance for respondent with the highest number of days traveled>

Potential Applications 3

Examining the first law of geography in a person level

(1) ECDF(Empirical Cumulative Distribution Function)



<ECDF of travel distance for respondents who belongs to Group 5 (n = 100)>

4. Summary & Discussion

Summary & Discussion

- Examined the potential for applying and utilizing a new type of smartphone based VGI dataset in the study of travel behavior(human mobility) analysis
- Time geography-based data pre-processing approach helps to Identify the five issues associated with the VGI dataset, including incomplete records, same time departure and arrival, nested trips, extremely low or high speed, and regularity
- The results of the case study suggest the smartphone-based VGI dataset may offer valuable insights into the longitudinal analysis of the impact of COVID-19 social distancing policies in Korea
- The land-use characteristics vary significantly at different times of the day, and these variations play a crucial role in utilizing the X,Y coordinate-timestamp only Volunteered Geographic Information (VGI) dataset for activity-travel behavior analysis
- To address regularity issues, it is beneficial to aggregate the data at a monthly level, even when considering individuals who have volunteered to provide their location data for approximately over 70% of the data collection period

References

- Brzinsky-Fay, C., Ebner, C., & Seibert, H. (2016). Veränderte Kontinuität. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 68(2), 229.
- Kim J. Y., Song T. J., Lee H. S., Kim D. H. (2018), Mobility Report, the Korea Transport Institute.
- Heiler, G., Reisch, T., Hurt, J., Forghani, M., Omani, A., Hanbury, A., & Karimipour, F. (2020, December). Country-wide mobility changes observed using mobile phone data during COVID-19 pandemic. In 2020 IEEE international conference on big data (big data) (pp. 3123-3132). IEEE.
- Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PloS one, 15(11), e0241957.
- Lee, J. H. (2016). Travel Behavior Dynamics in Space and Time (Doctoral dissertation, University of California, Santa Barbara).
- Lee, J. H., Davis, A. W., Yoon, S. Y., & Goulias, K. G. (2016). Activity space estimation with longitudinal observations of social media data. Transportation, 43, 955-977.
- Su, R., McBride, E. C., & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. Transportation Research Part C: Emerging Technologies, 120, 102796.
- Wang, X., Pei, T., Li, K., Cen, Y., Shi, M., Zhuo, X., & Mao, T. (2022). Analysis of changes in population's cross-city travel patterns in the pre-and post-pandemic era: A case study of China. Cities, 122, 103472.
- White, C. E., Bernstein, D., & Kornhauser, A. L. (2000). Some map matching algorithms for personal navigation assistants. Transportation research part c: emerging technologies, 8(1-6), 91-108.
- Yoon, S., & Goulias, K. (2010). Impact of time-space prism accessibility on time use behavior and its propagation through intra-household interaction. Transportation Letters, 2(4), 245-260.
- Yoon, S., Lee, J.H., Kong, J.H. (2022), Analyzing the effect of social-distancing policies based on the smartphone based VGI dataset, KRIHS

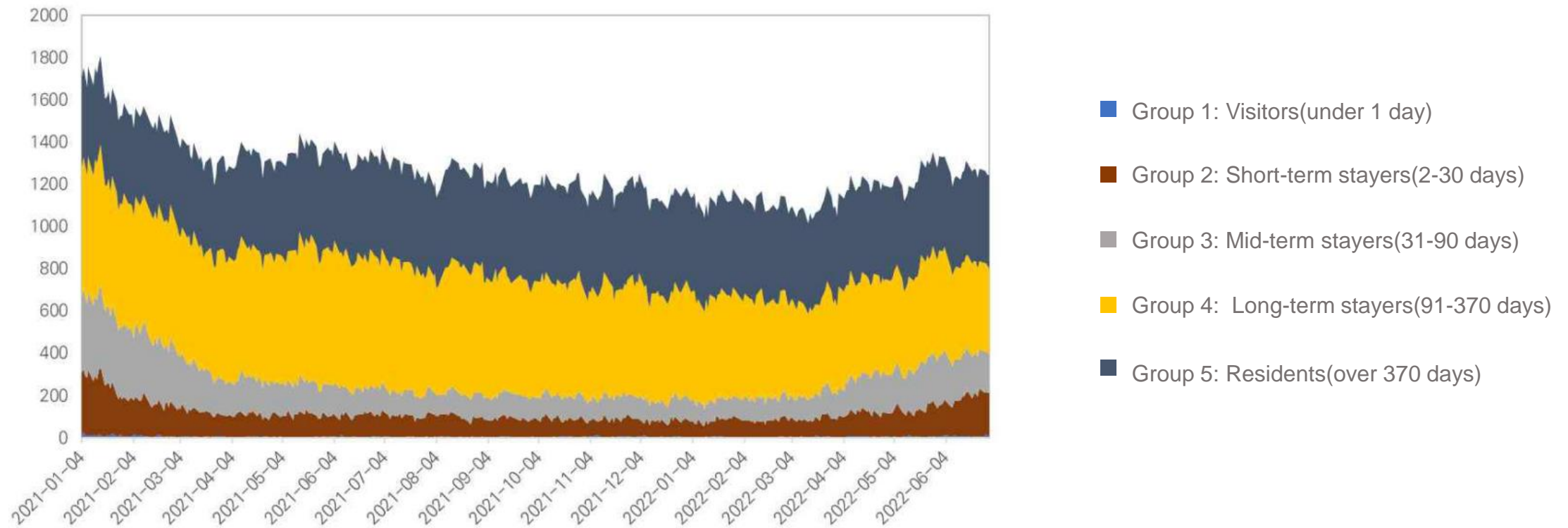
**Thank
you**



GIScience & Geospatial Bigdata Laboratory
Department of geography / Kyungpook National University

Potential Applications 1

Visitors vs Stayers vs Residents

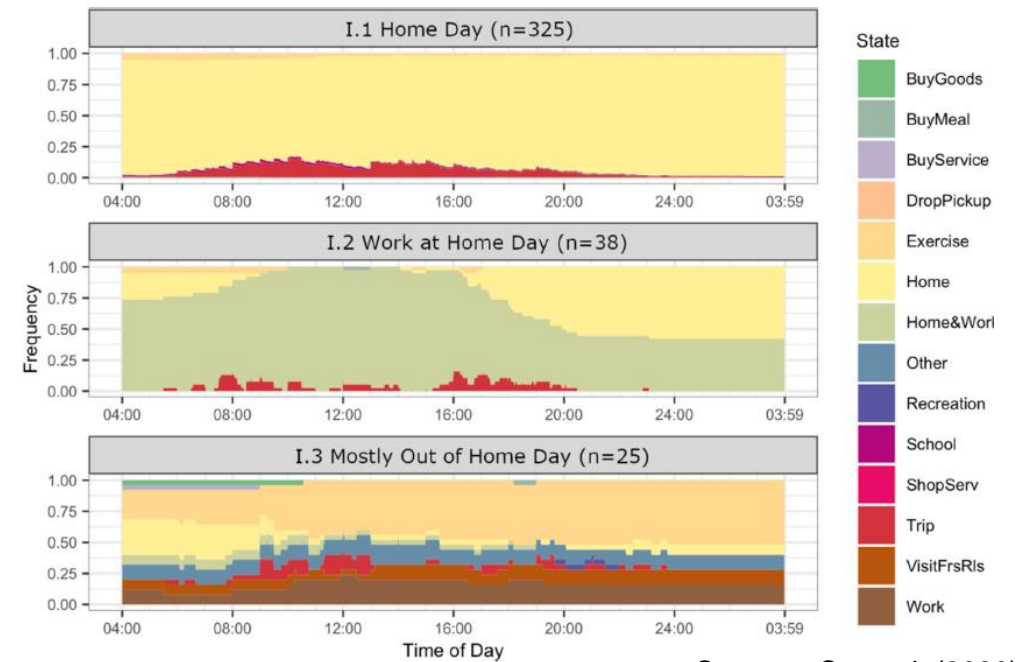
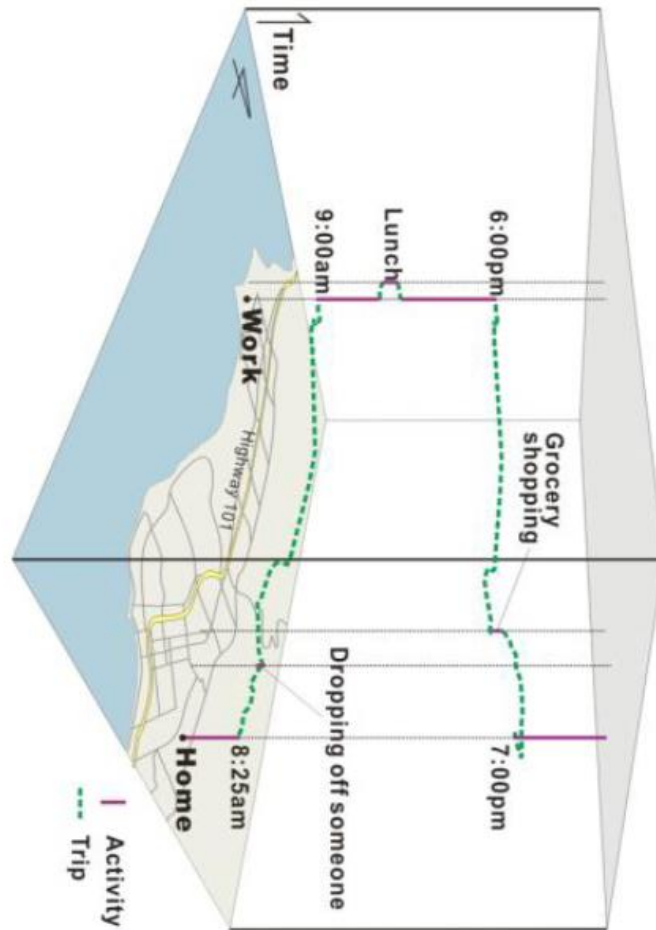


Sources: Yoon et al. (2022)

<Number of people observed within a day during the data collection period(excluding weekends data)>

Case study

Methodology

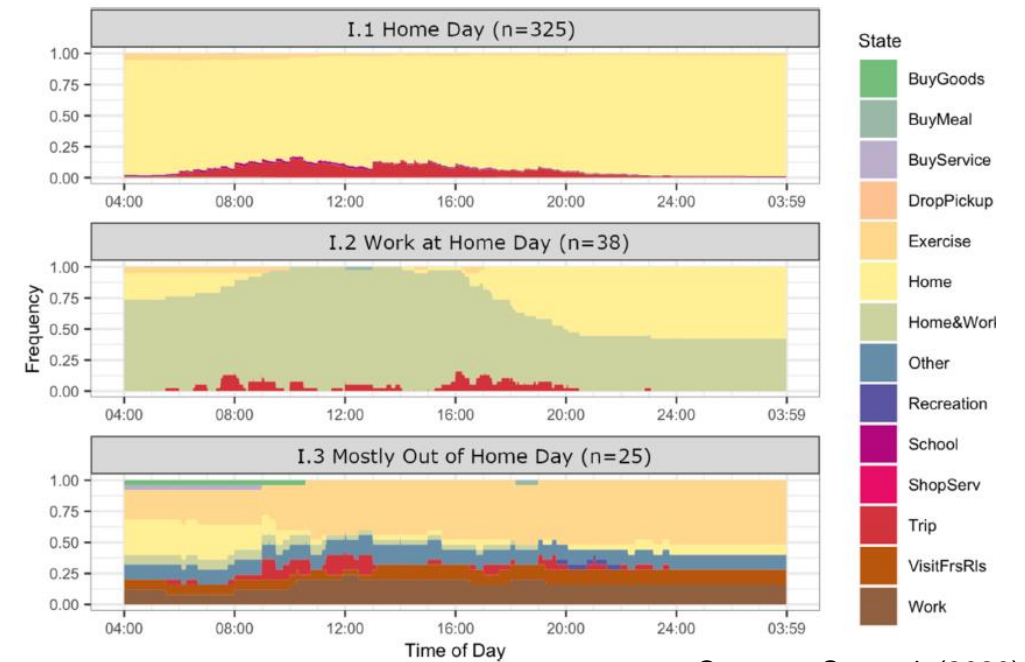
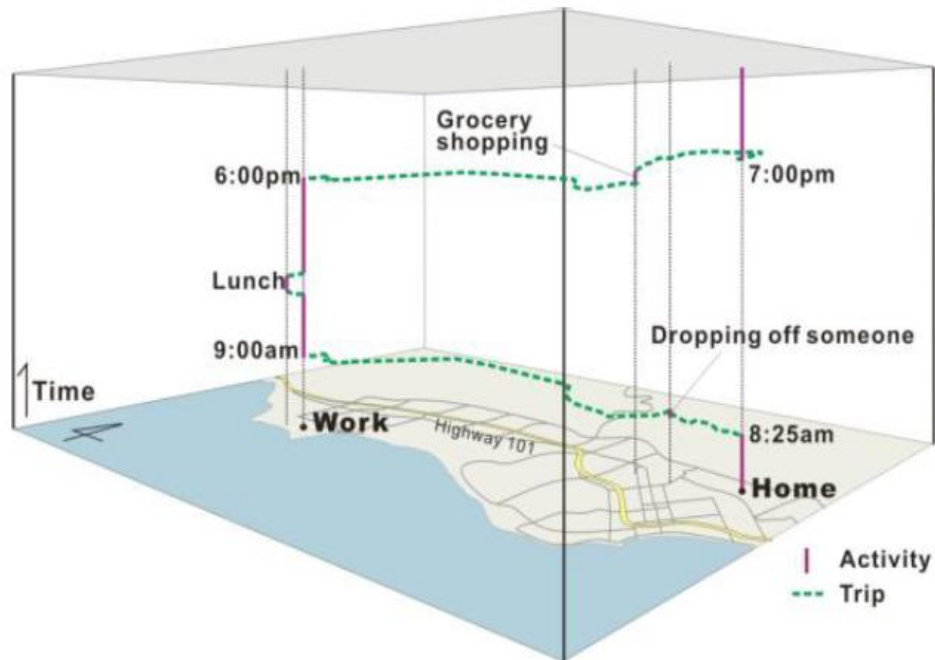


Sources: Su et al. (2020)

<Example – Daily time of day patterns of activity sequences>

Case study

Methodology



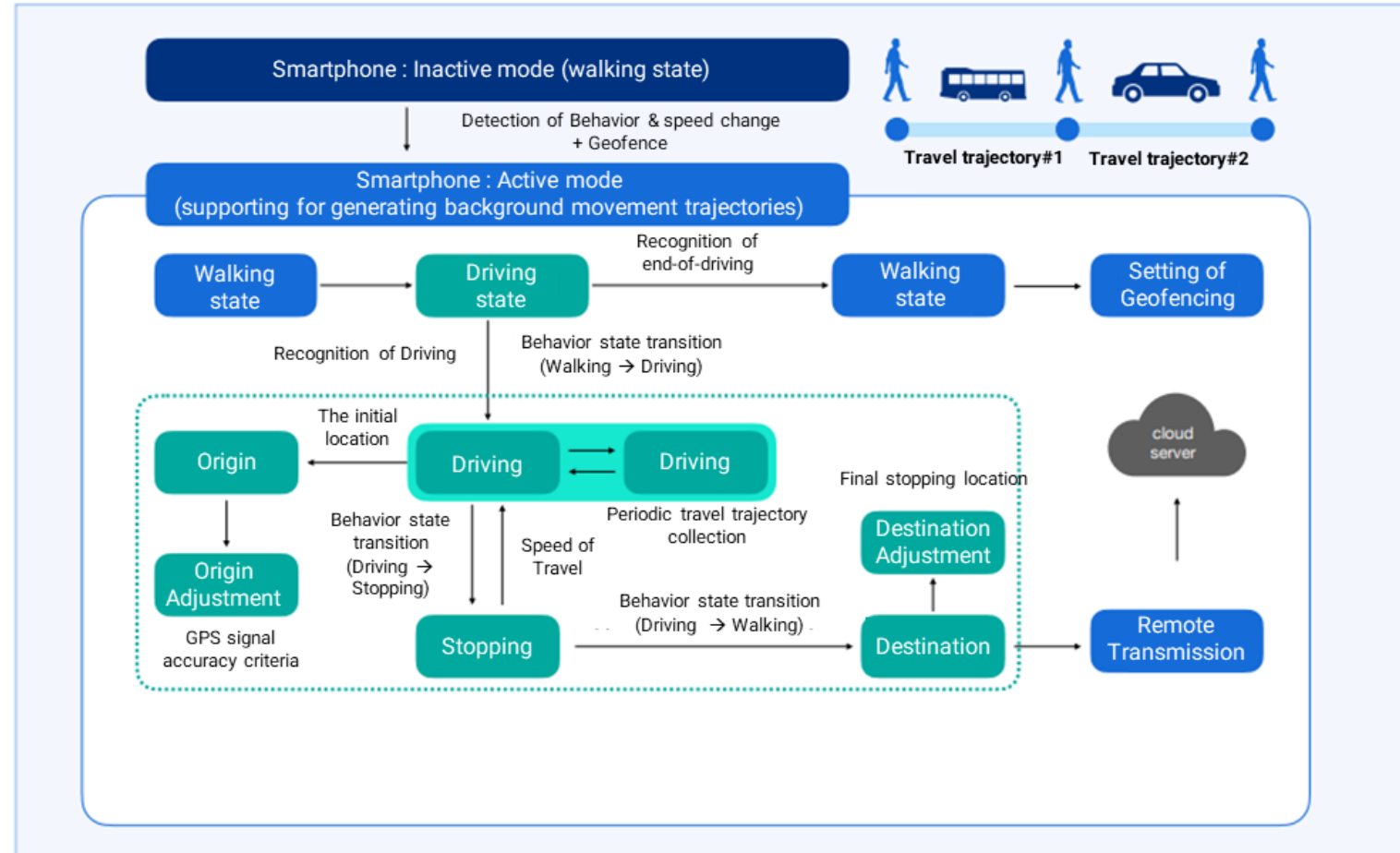
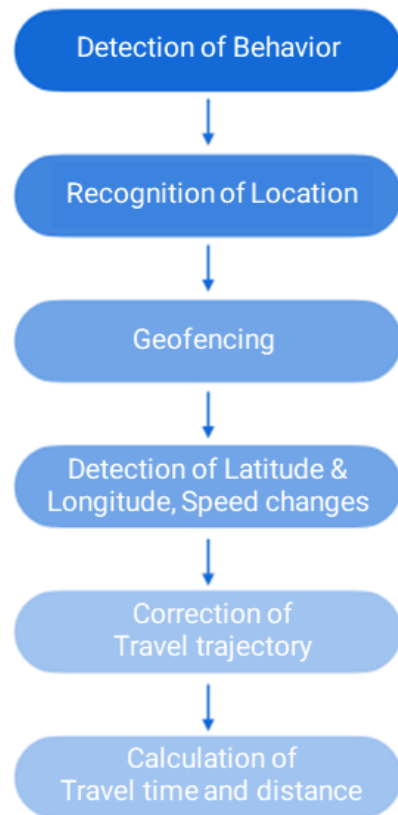
Sources: Su et al. (2020)

<Example – Daily time of day patterns of activity sequences>

Appendix

AI-based crowdsourced traffic information collection technology

Sourcing-driven information collection



Case study

| Objectives

- Identification of mobility patterns based on land use characteristics at activity locations and sequence analysis
- Impact analysis of social distancing policies using longitudinal VGI datasets

Case study

Results : Departure and Arrival

(4) Landuse characteristics of Departure location by different groups

Sources: Yoon et al. (2022)

Example)

Num of departure location

250

200

150

100

50

0

0:00 1:00 2:00 3:00 4:00 5:00 6:00 7:00 8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00

Num of departure location
(only 'Out of Daegu')

500

450

400

350

300

250

200

150

100

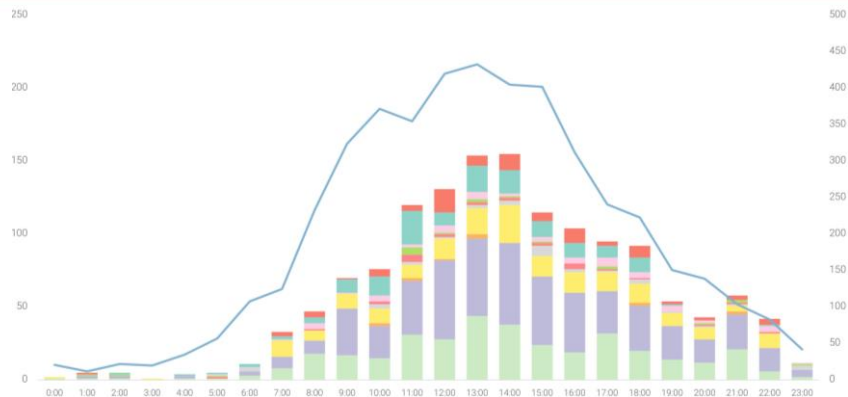
50

0

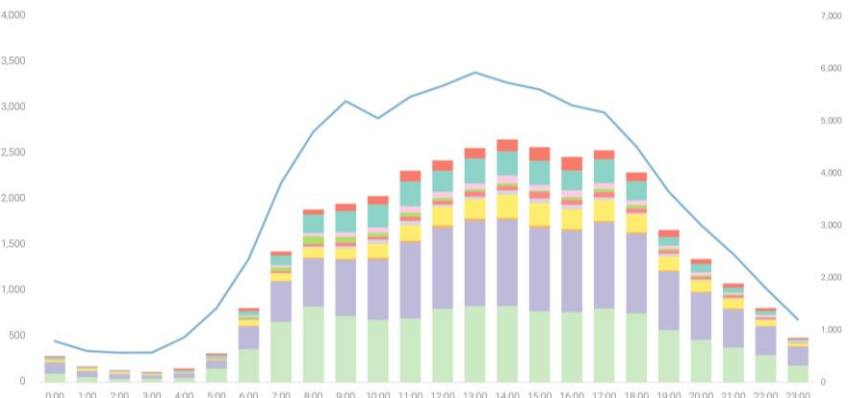
- Residential areas
- Neighborhood living facilities
- Cultural & Sports
- Commercial
- Medical
- Education & Research
- Hotels
- Business
- Factory & Garage
- Other
- Out of Daegu
- Traveling

Case study

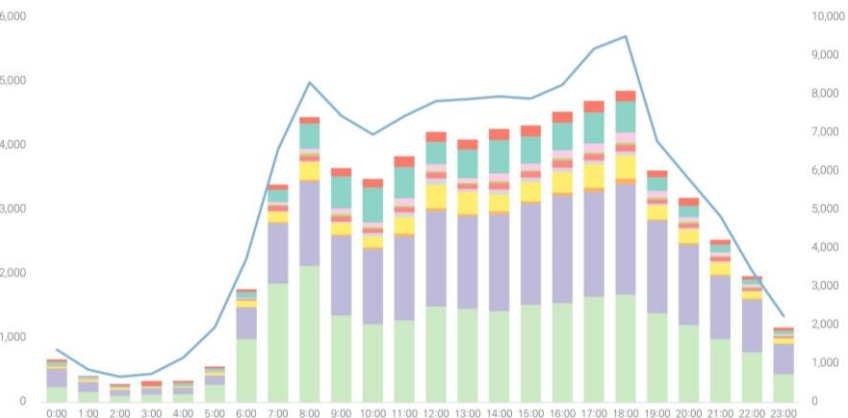
[Group 1]



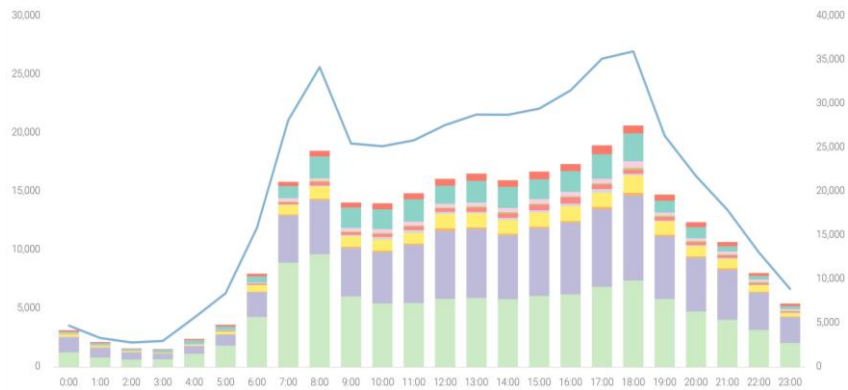
[Group 2]



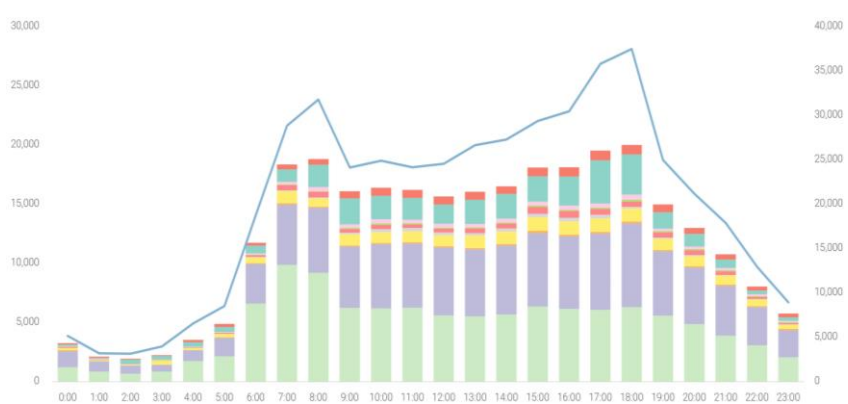
[Group 3]



[Group 4]



[Group 5]



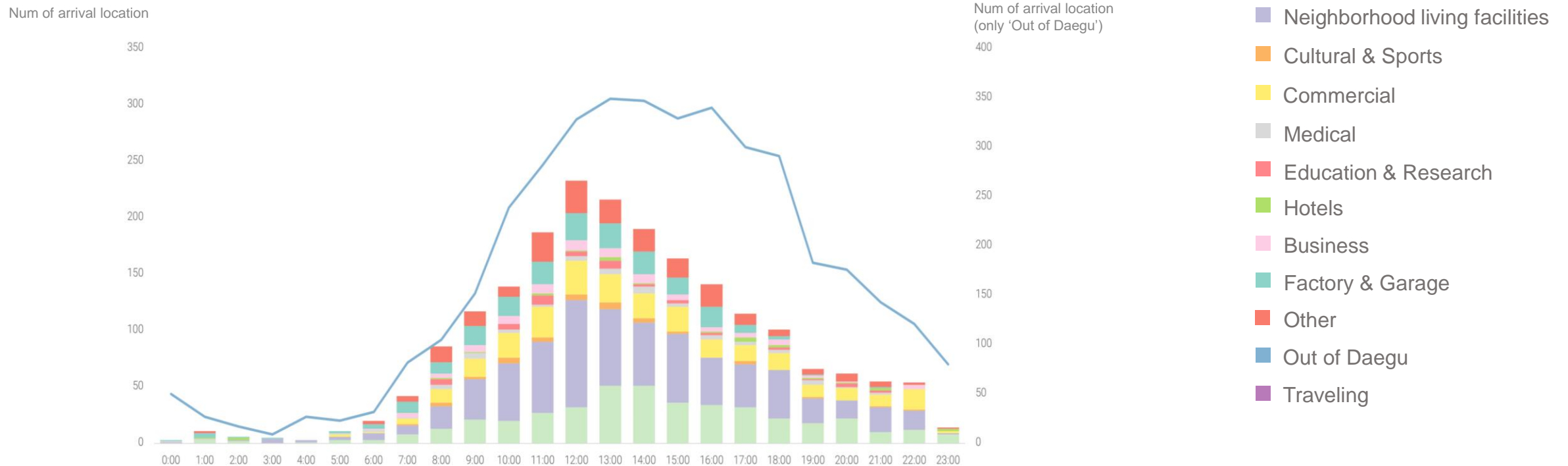
Case study

Results : Departure and Arrival

(5) Landuse characteristics of Arrival location by different groups

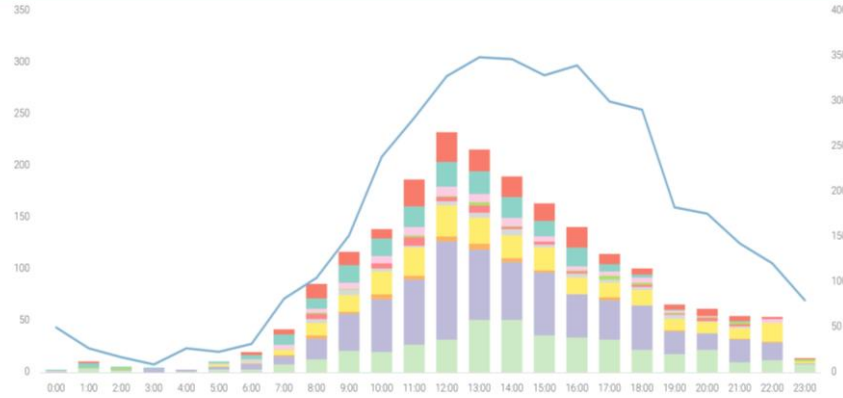
Sources: Yoon et al. (2022)

Example)

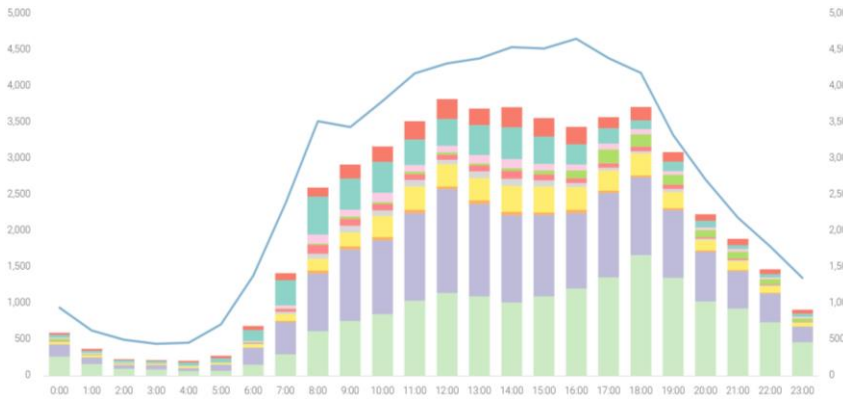


Case study

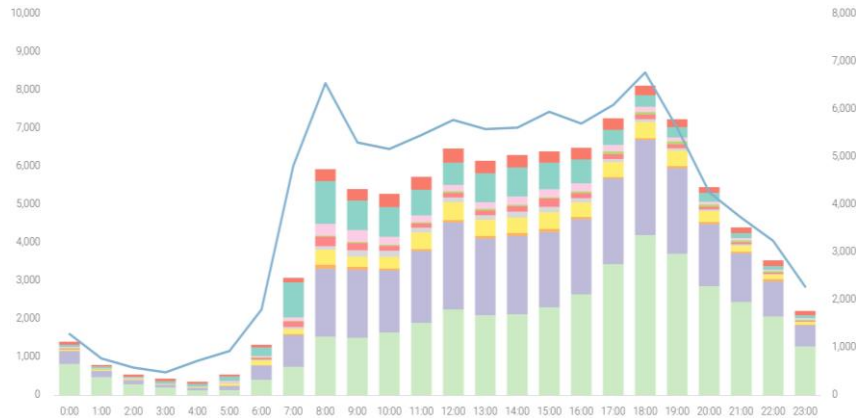
[Group 1]



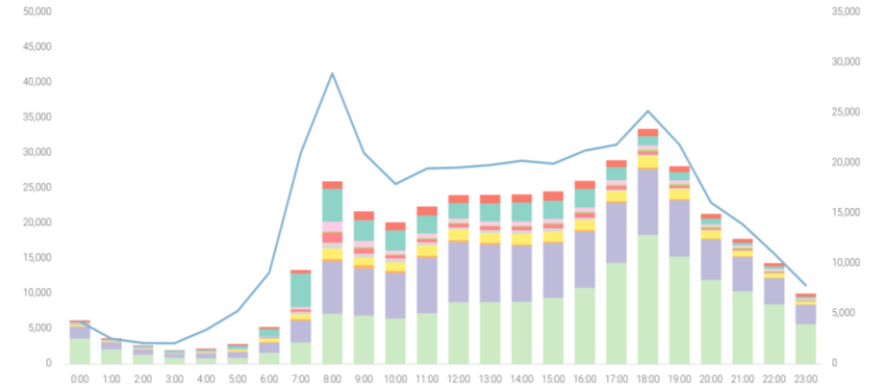
[Group 2]



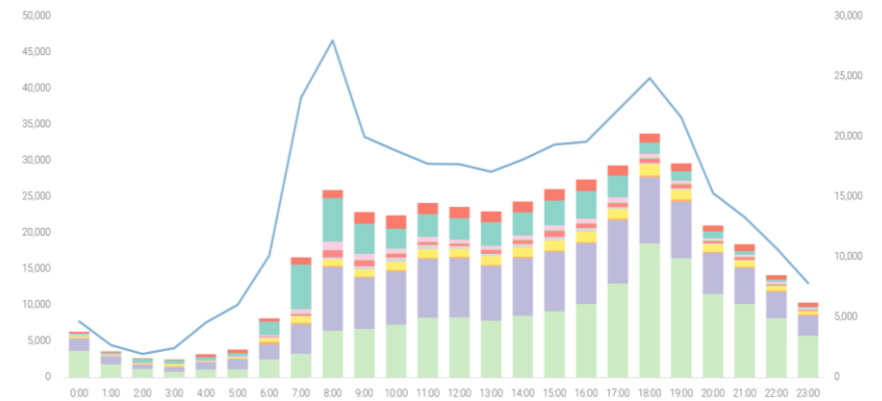
[Group 3]



[Group 4]



[Group 5]



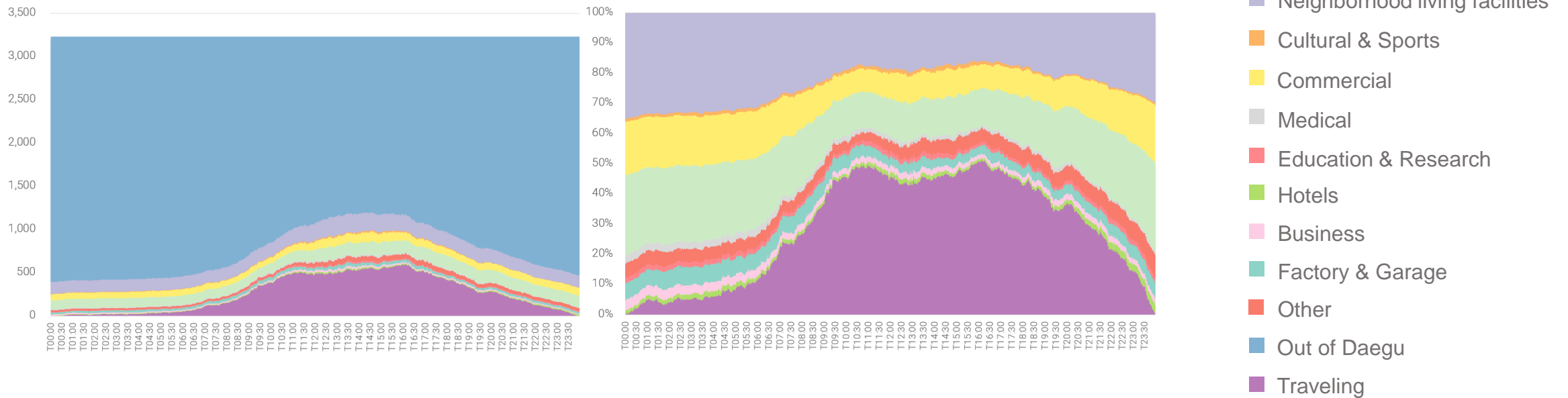
Case study

Results : Activity location logs

(1) Basic analysis of activity location logs (left : N, right : Proportion)

Sources: Yoon et al. (2022)

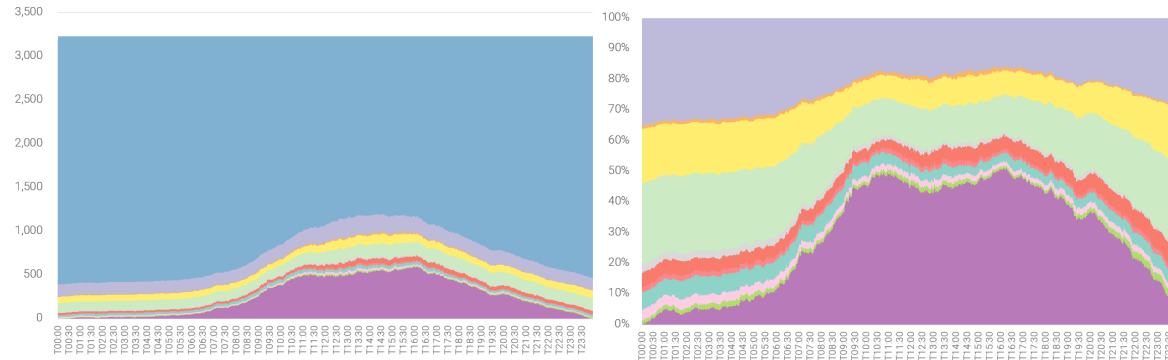
Example)



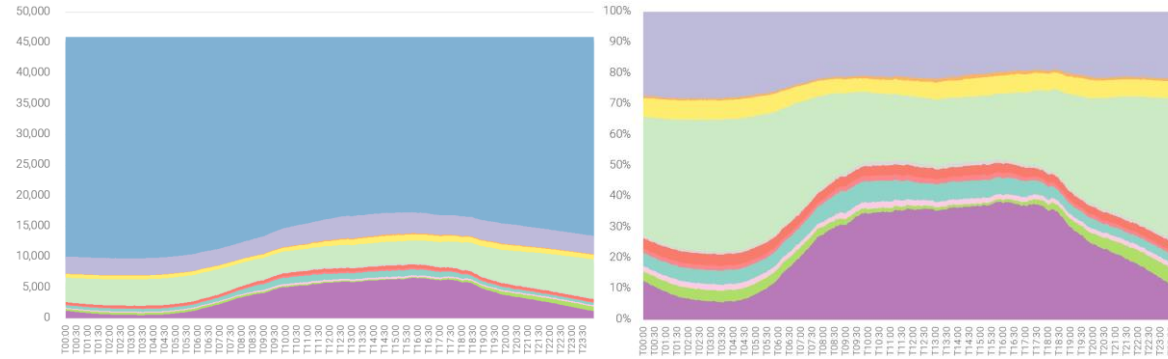
<Basic analysis of activity location logs by different groups>

Case study

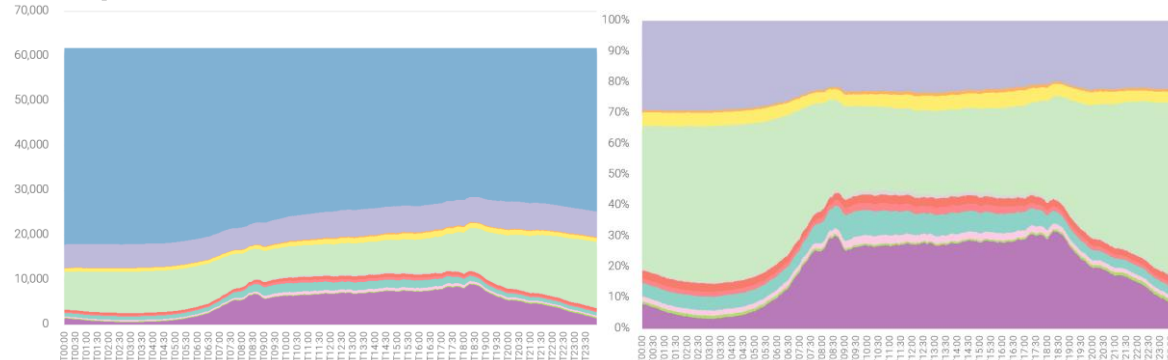
[Group 1]



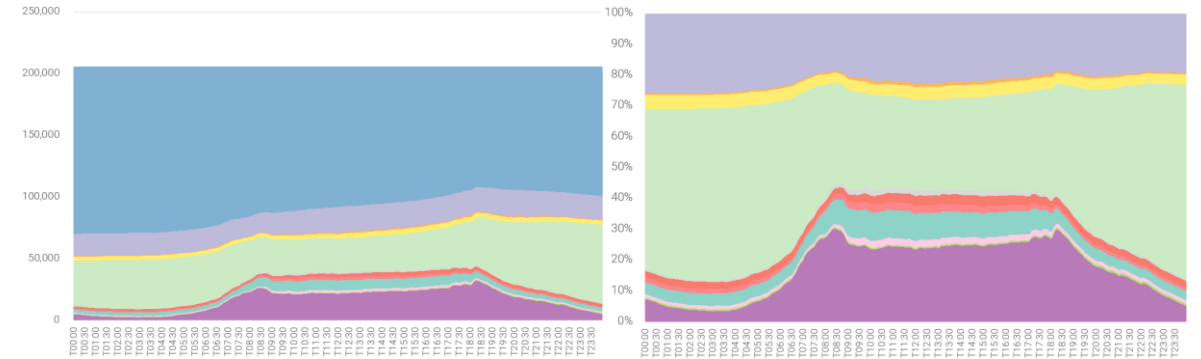
[Group 2]



[Group 3]



[Group 4]



[Group 5]

