UNIVERSITY OF ARKANSAS

SafeGragh Social Distancing Metrics for COVID-19 Studies

Xiao Huang – xh010@uark.edu

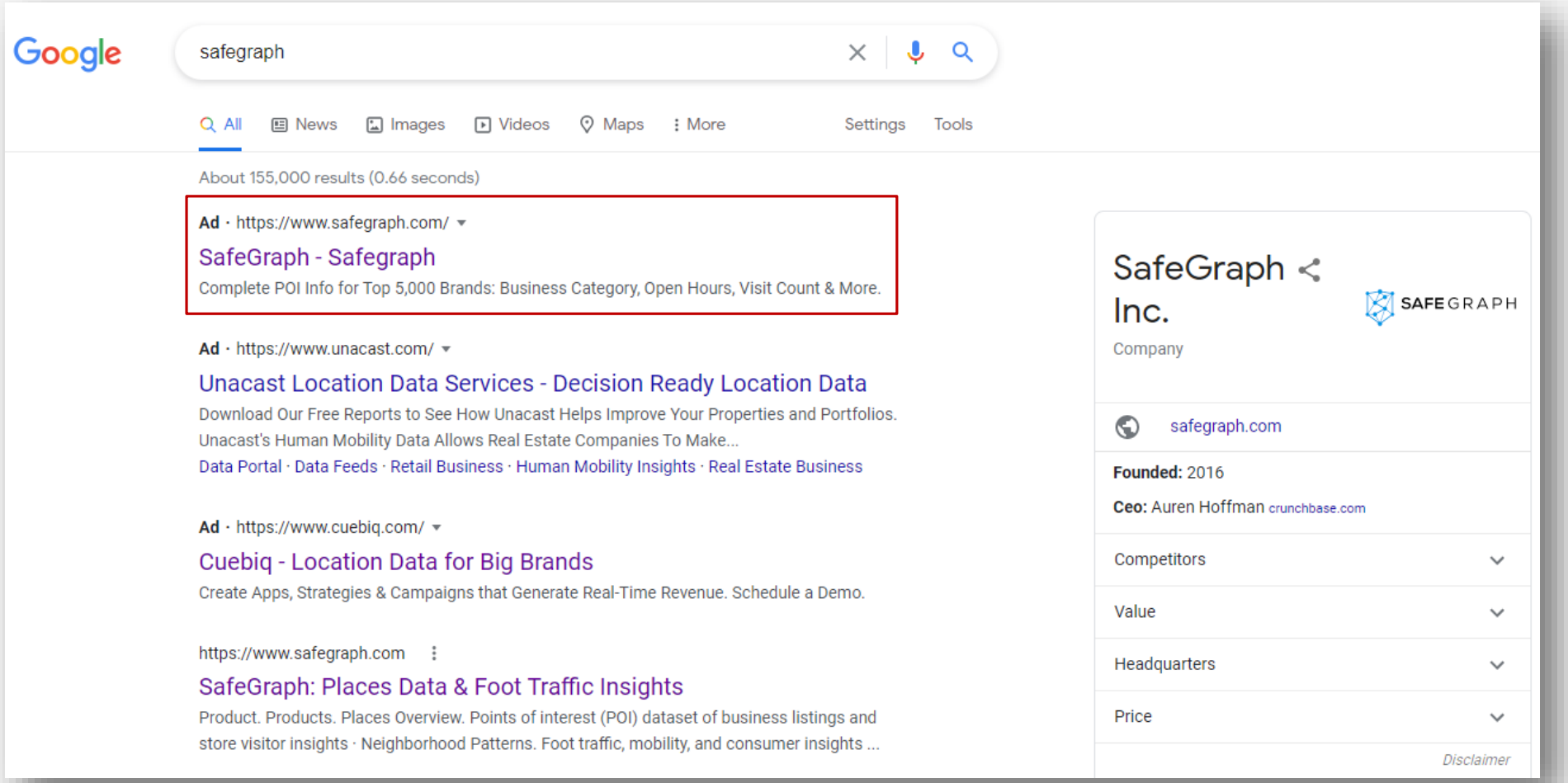
May 07, 2021

UNIVERSITY OF ARKANSAS

SafeGragh Social Distancing Metrics for COVID-19 Studies

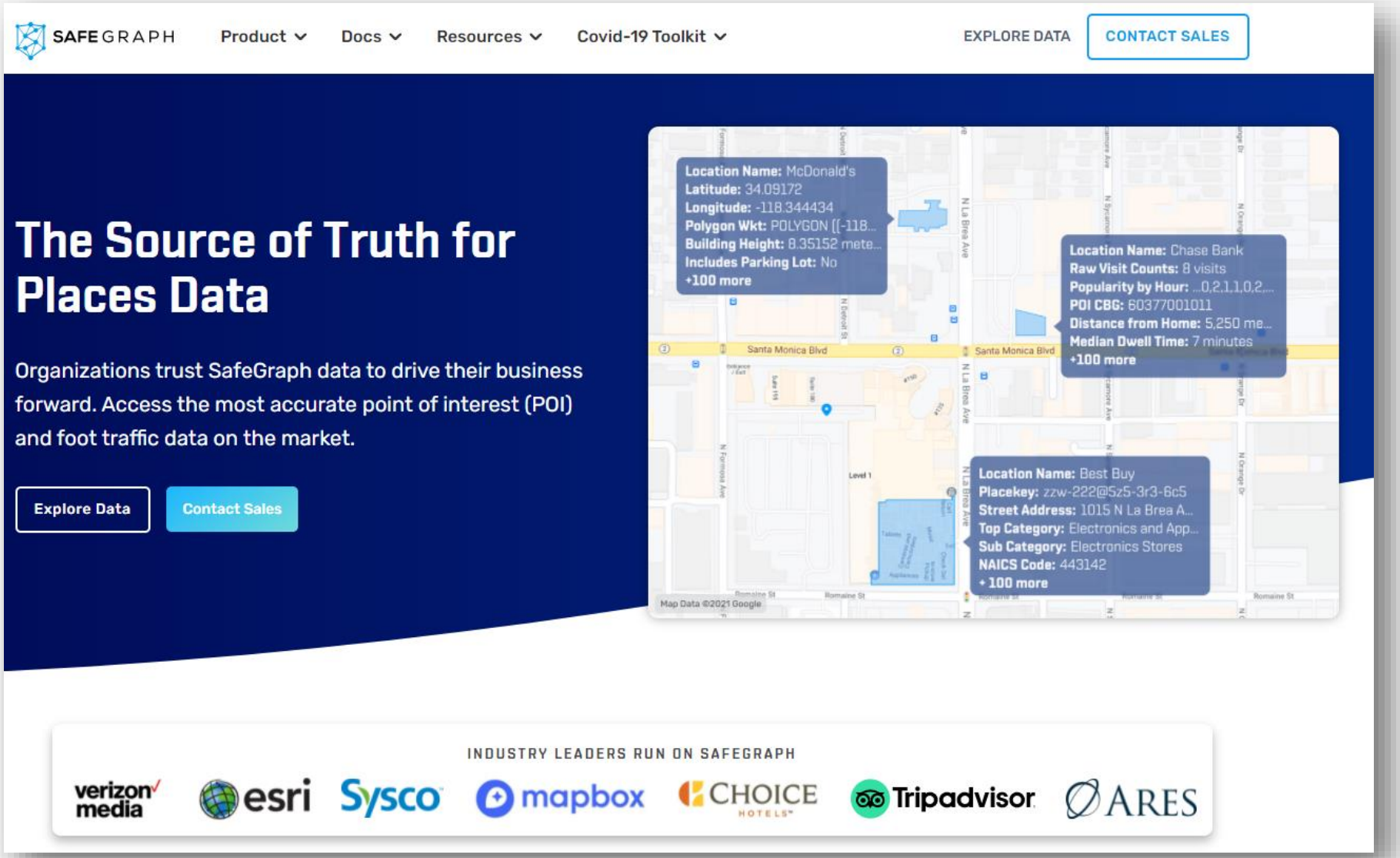
- How to obtain SafeGraph COVID-19 data?
- How Social Distancing Metrics are derived?
- What information do Social Distancing Metrics contain?
- What do Social Distancing Metrics look like?
- How can these data be utilized in COVID-19 studies

What is SafeGraph?



SAFEGRAPH

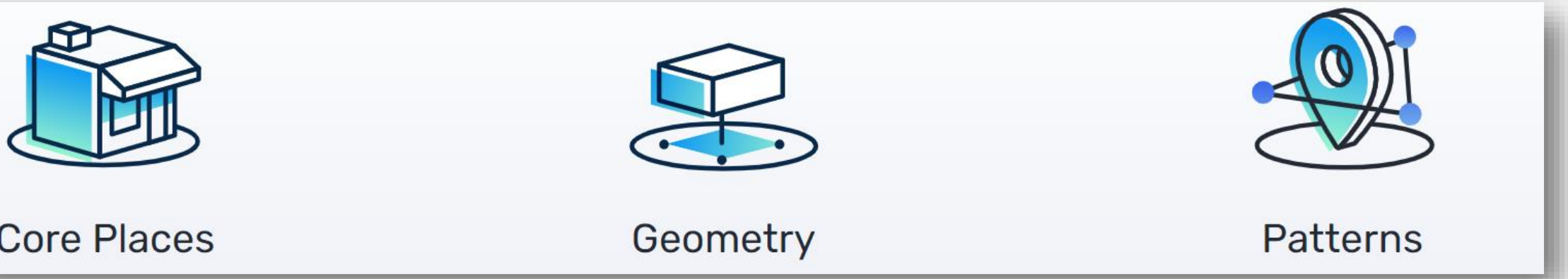
What is SafeGraph?



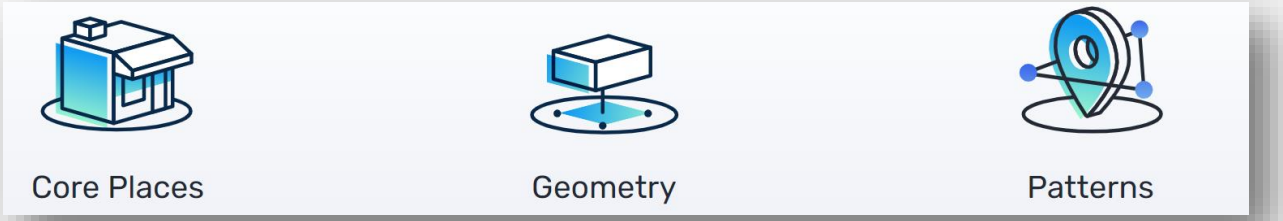
SAFEGRAPH

A commercial company that aims to provide **insights** about physical places by **aggregating anonymized location data** from numerous applications and devices.

What is SafeGraph?



What is SafeGraph?

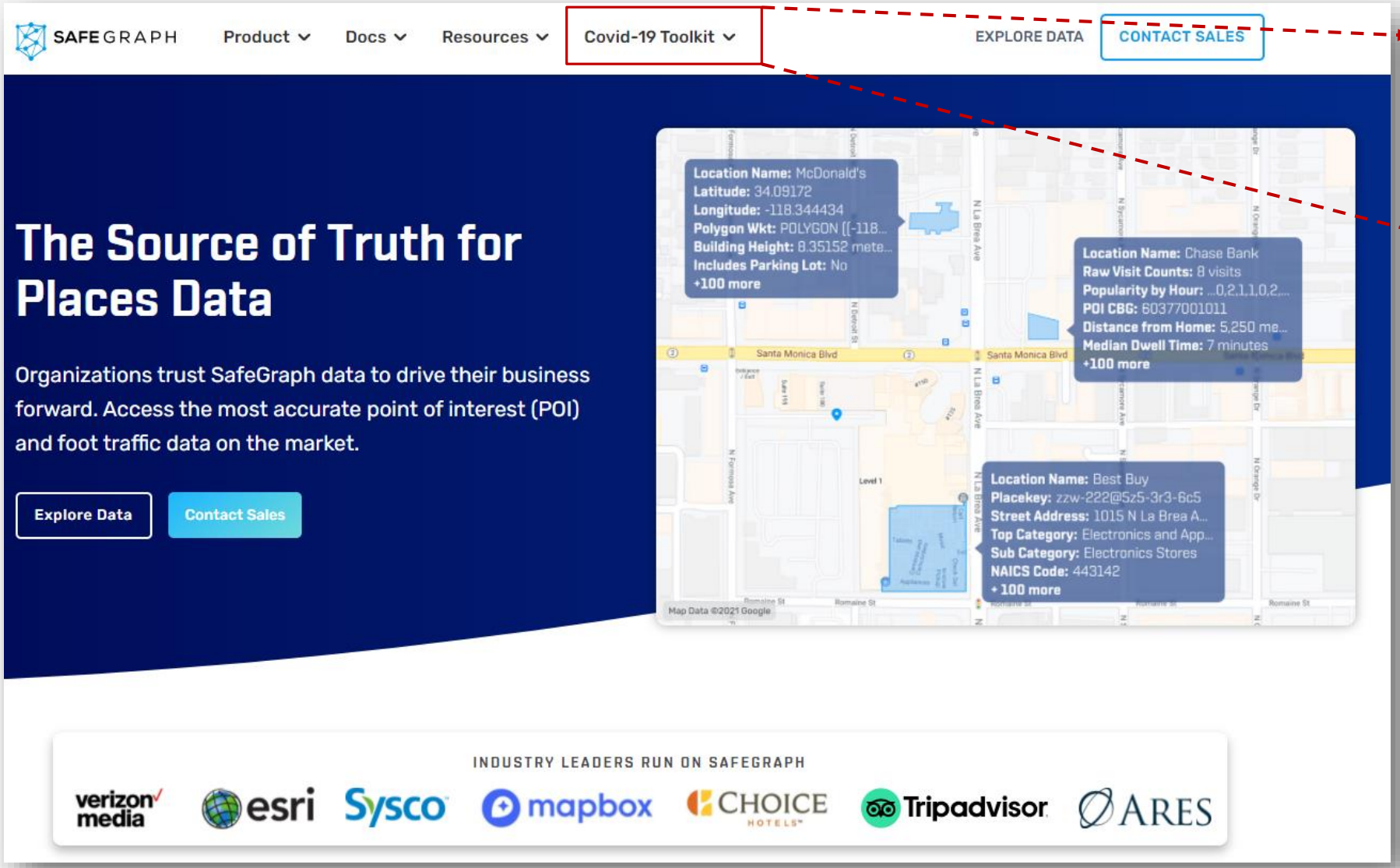


- ❖ Core Places: Base information such as location name, address, category, and brand association for points of interest (POIs) where people spend time or money. Available for ~8.4M POI including permanently closed POIs.

- ❖ Geometry: POI footprints with spatial hierarchy metadata depicting when child polygons are contained by parents or when two tenants share the same polygon. Available for ~7.8M POI (Geometry metadata not provided for closed POIs)

- ❖ Patterns: Place traffic and demographic aggregations that answer: how often people visit, how long they stay, where they came from, where else they go, and more. Available for ~4.5M POIs.

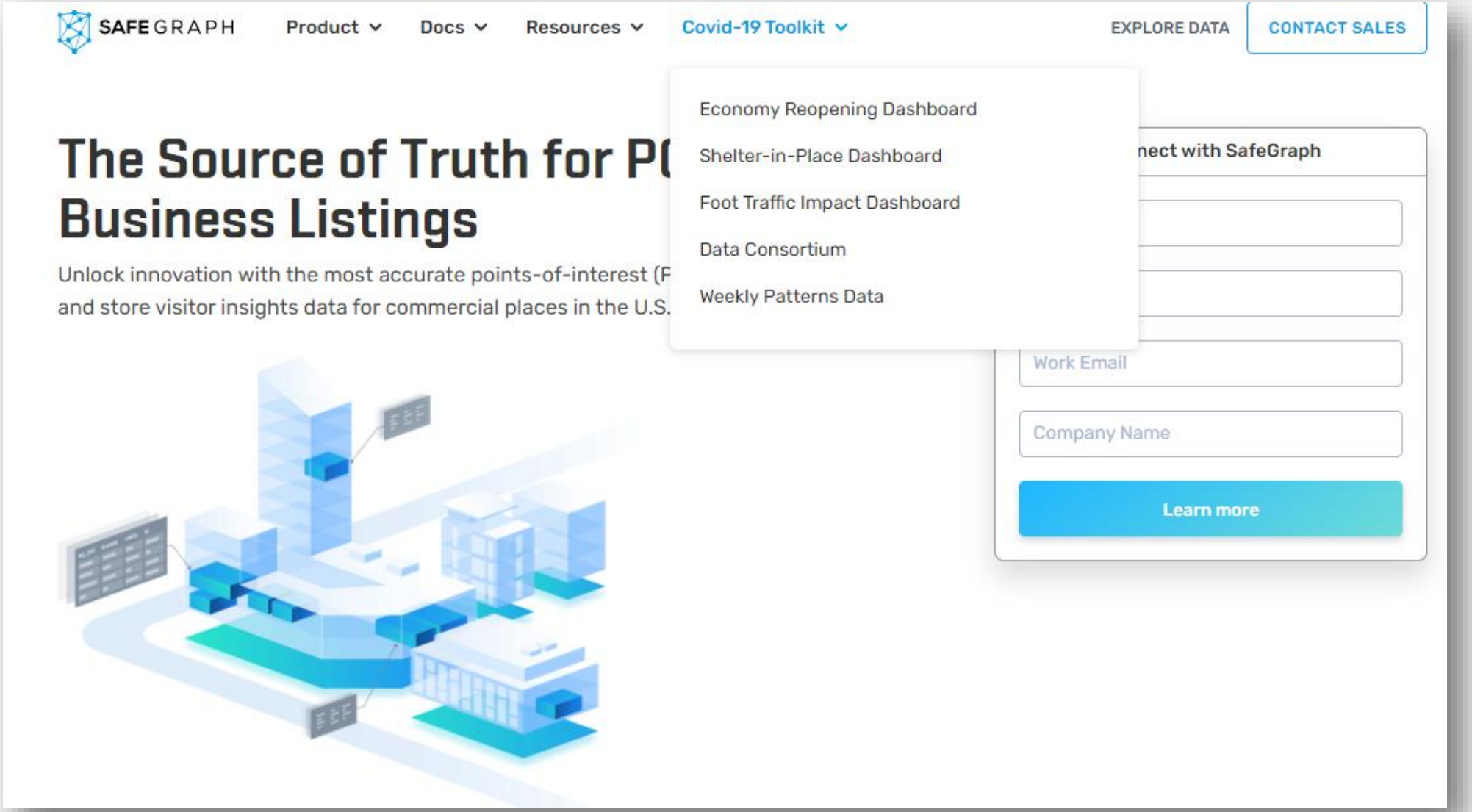
How to obtain SafeGraph COVID-19 data?



COVID-19 Toolkit

SAFEGRAPH

How to obtain SafeGraph COVID-19 data?



January 3 — December 27 | All Industries | All States

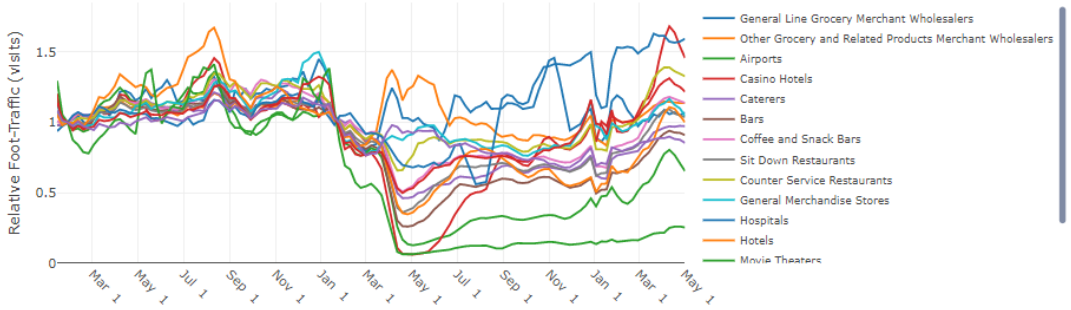
Selected Date | Foot Traffic by State | Counter Service Restaurants | Sit Down Restaurants

February 7 — September 13 | Shelter in Place Index | All States | County-level

Shelter in Place Index by Region

INDUSTRY LEVEL PATTERNS

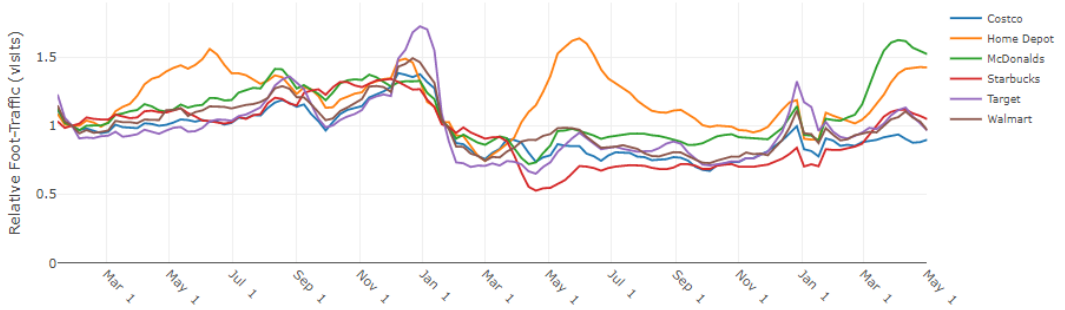
All Industries



Note: Click an industry in the legend to toggle it on/off

BRAND SPECIFIC PATTERNS

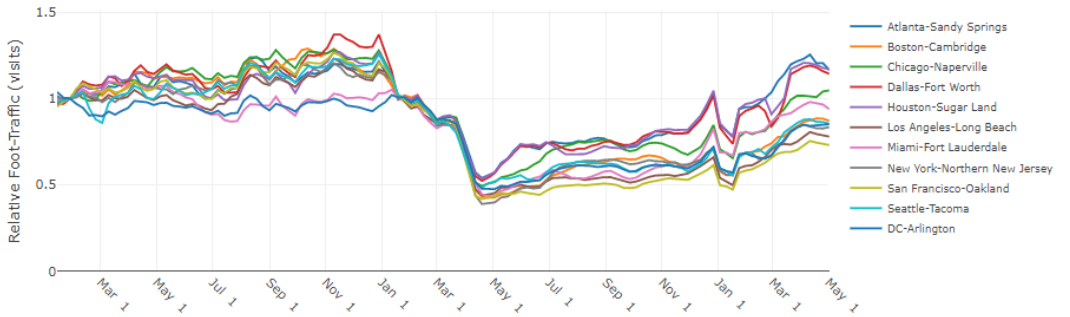
All Brands



Note: Click a brand in the legend to toggle it on/off

TOP IMPACTED REGIONS

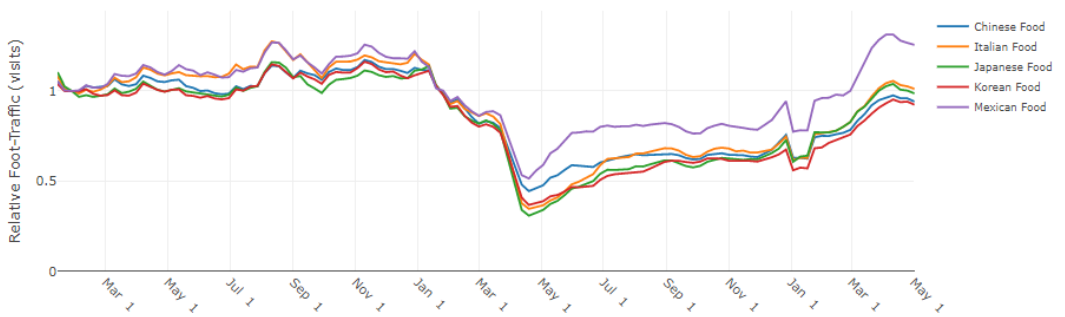
All Regions



Note: Click a region in the legend to toggle it on/off

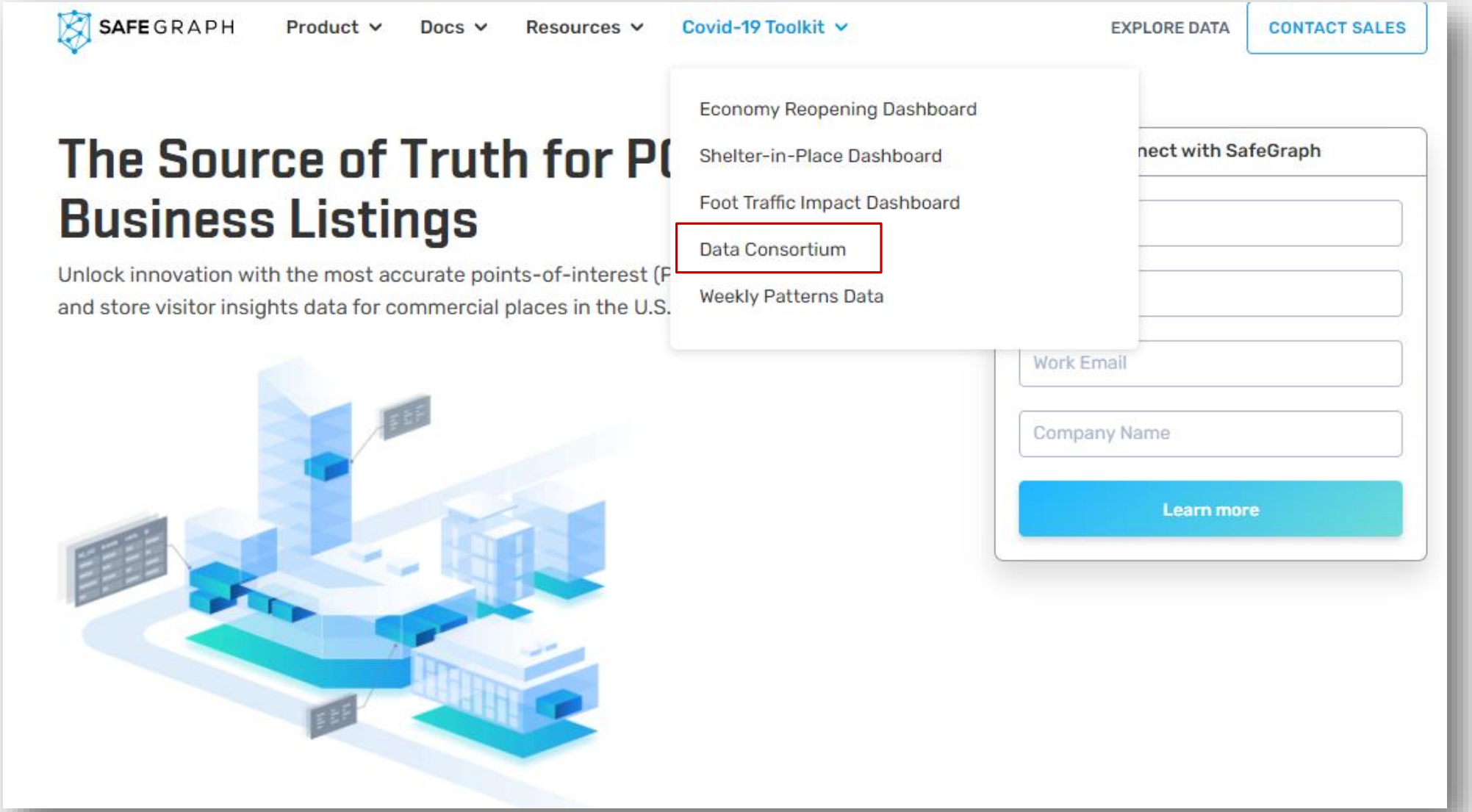
RESTAURANT CATEGORIES

All Restaurant Categories

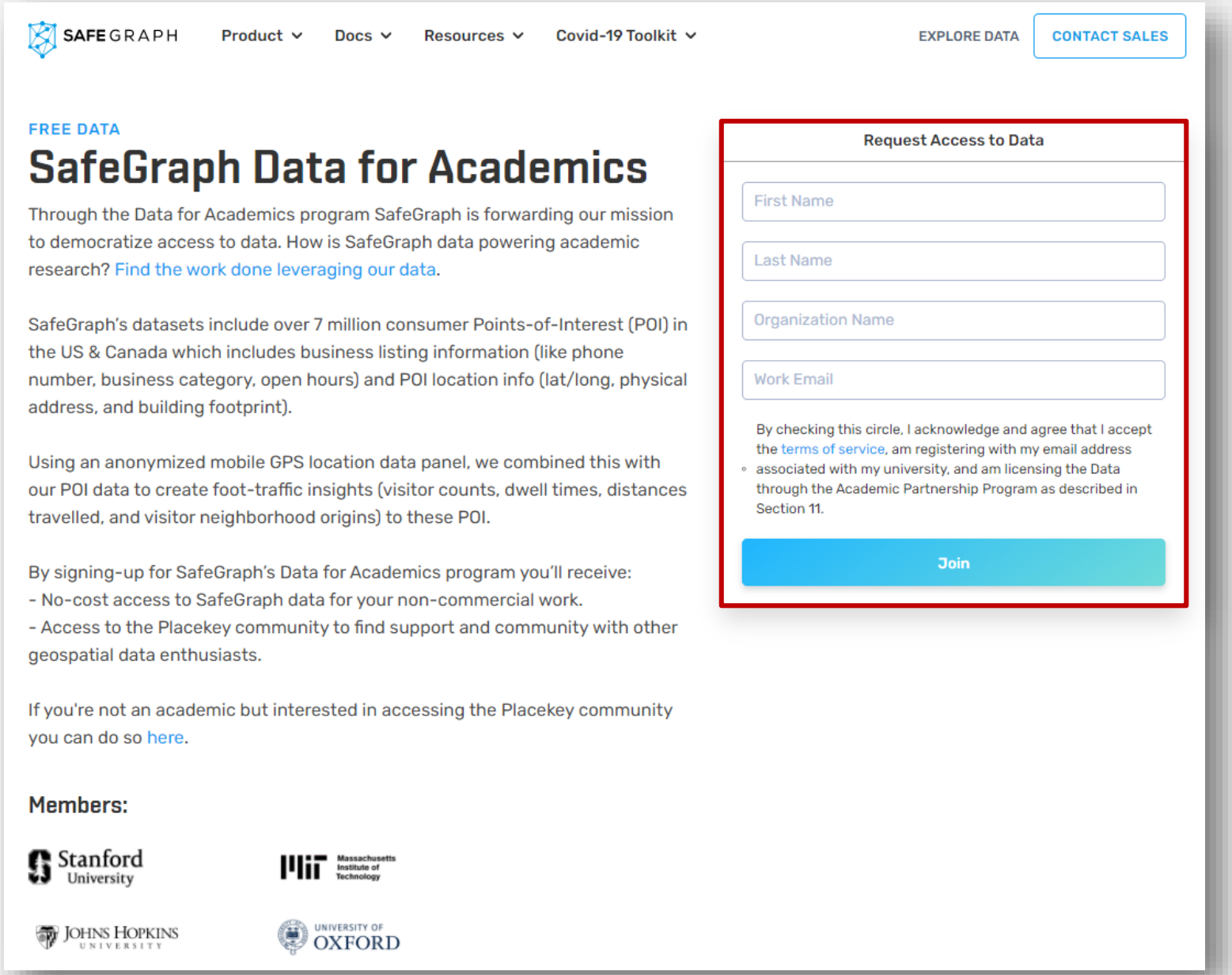


Note: Click a restaurant category in the legend to toggle it on/off

How to obtain SafeGraph COVID-19 data?



How to obtain SafeGraph COVID-19 data?



How to obtain SafeGraph COVID-19 data?

SAFEGRAPH Product Docs Resources Covid-19 Toolkit EXPLORE DATA CONTACT SALES

FREE DATA

SafeGraph Data for Academics

Through the Data for Academics program SafeGraph is forwarding our mission to democratize access to data. How is SafeGraph data powering academic research? Find the work done leveraging our data.

SafeGraph's datasets include over 7 million consumer Points-of-Interest (POI) in the US & Canada which includes business listing information (like phone number, business category, open hours) and POI location info (lat/long, physical address, and building footprint).

Using an anonymized mobile GPS location data panel, we combined this with our POI data to create foot-traffic insights (visitor counts, dwell times, distances travelled, and visitor neighborhood origins) to these POI.

By signing-up for SafeGraph's Data for Academics program you'll receive:
- No-cost access to SafeGraph data for your non-commercial work.
- Access to the Placekey community to find support and community with other geospatial data enthusiasts.

If you're not an academic but interested in accessing the Placekey community you can do so here.

Members:

Stanford University · Massachusetts Institute of Technology · Johns Hopkins University · University of Oxford

Request Access to Data

First Name

Last Name

Organization Name

Work Email

By checking this circle, I acknowledge and agree that I accept the terms of service, am registering with my email address associated with my university, and am licensing the Data through the Academic Partnership Program as described in Section 11.

Join

Request Access to Data

Xiao

Huang

University of Arkansas

xh010@uark.edu

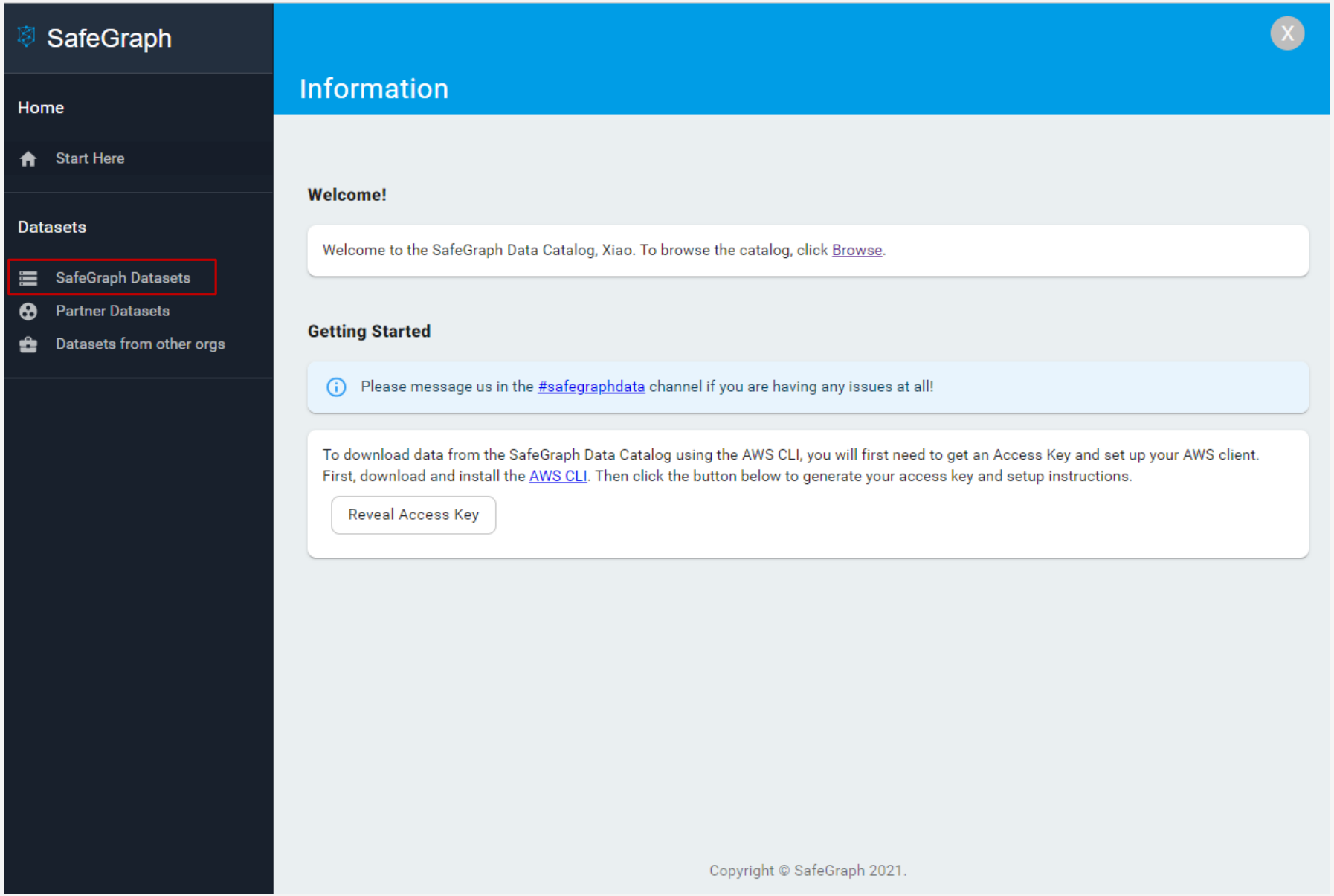
By checking this circle, I acknowledge and agree that I accept the terms of service, am registering with my email address associated with my university, and am licensing the Data through the Academic Partnership Program as described in Section 11.

Join

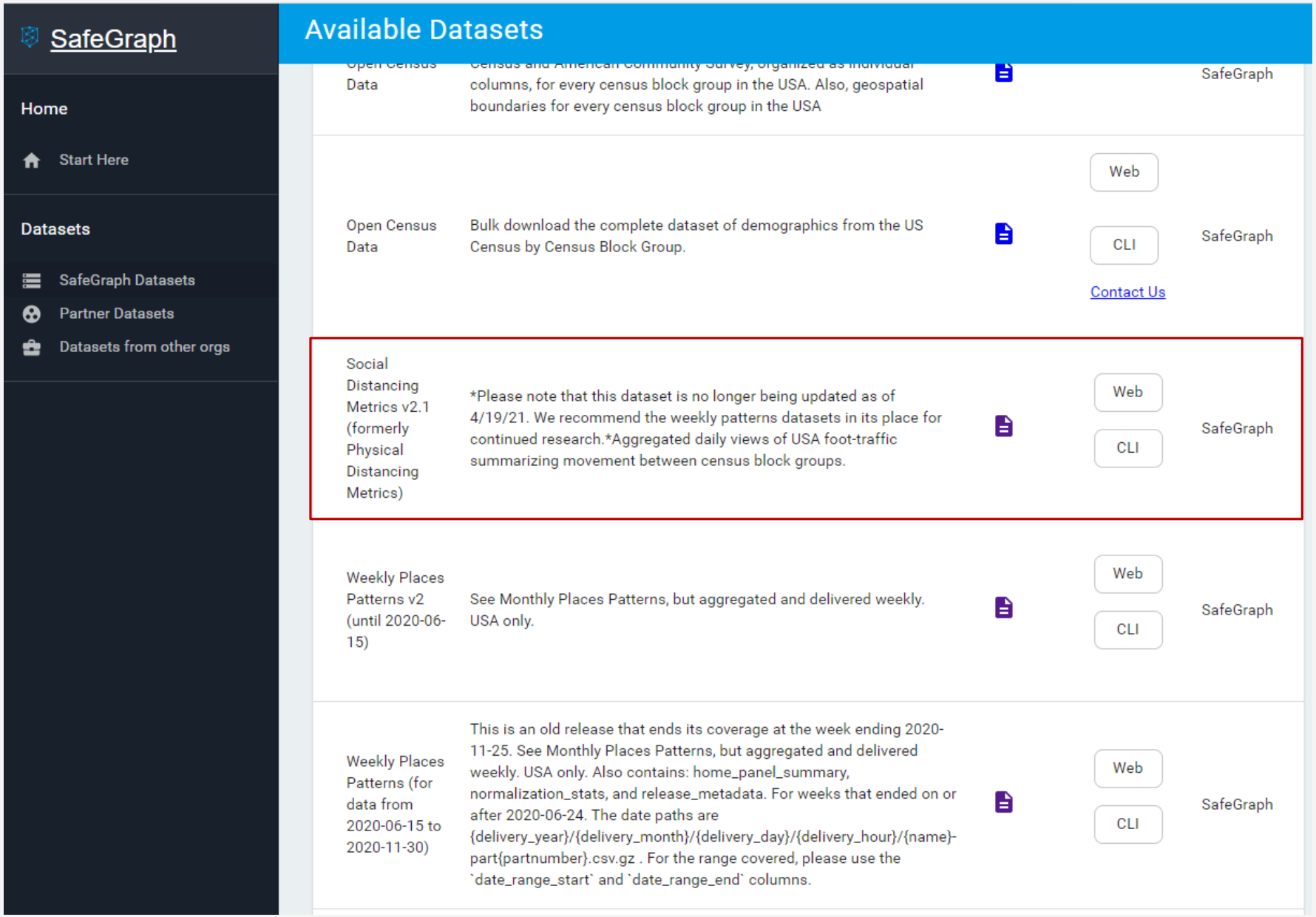
How to obtain SafeGraph COVID-19 data?

Wait for an email…

How to obtain Social Distancing Metrics?



How to obtain Social Distancing Metrics?



Note: updated as of 04/19/2021

01/01/2019

04/19/2021

How Social Distancing Metrics are derived?

- "The data was generated using a panel of **GPS pings** from 45 million **anonymous mobile devices"**

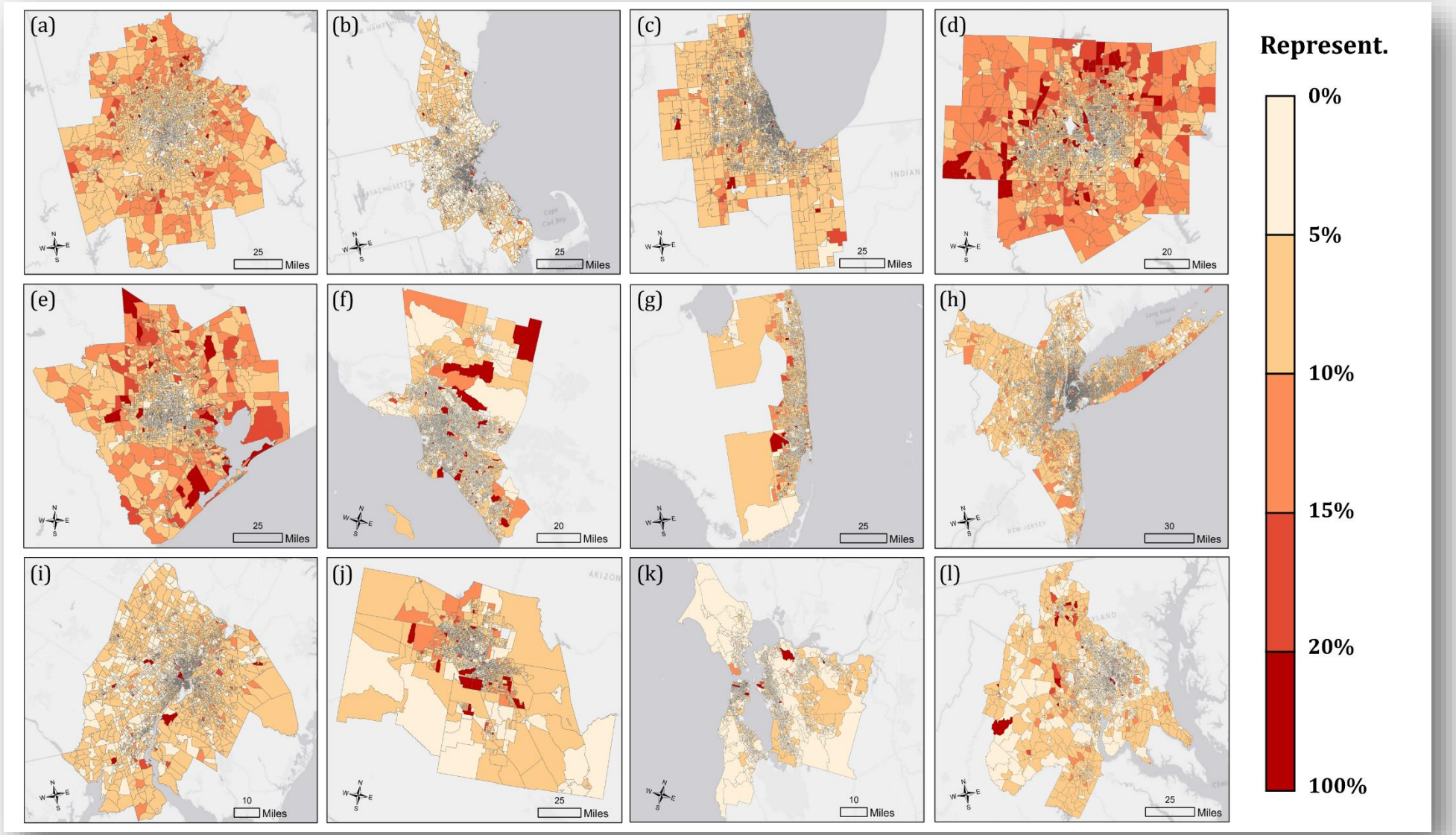
- "We determine the common nighttime location (**devices' home**) of each mobile device over a **6-week** period to a Geohash-7 granularity (~153m x ~153m)"

- We then aggregate the devices by **home census block group** and provide the metrics set out below for each census block group.

Representativeness?

"Although SafeGraph Patterns aggregates data from **~ 10%** of devices in the United States, this sample **IS NOT** a perfect representative subset of the population"

Representativeness?



(a) Atlanta MSA;

(b) Boston MSA;

(c) Chicago MSA

(d) Dallas MSA;

(e) Houston MSA;

(f) Los Angeles MSA;

(g) Miami MSA;

(h) New York MSA;

(i) Philadelphia MSA;

(j) Phoenix MSA;

(k) San Francisco MSA;

(l) D.C. MSA.

What information do Social Distancing Metrics contain?

Column Name:	Descriptions:	Example:
origin_census_block_group	The unique **12-digit** FIPS code for the Census Block Group. Please note that some CBGs have leading zeros.	131000000000
date_range_start	Start time for measurement period in ISO 8601 format of YYYY-MM-DDTHH:mm:SS±hh:mm (local time with offset from GMT).	2020-03-01T00:00:00-06:00
date_range_end	End time for measurement period in ISO 8601 format of YYYY-MM-DDTHH:mm:SS±hh:mm (local time with offset from GMT).	2020-03-02T00:00:00-06:00

What information do Social Distancing Metrics contain?

Column Name:	Descriptions:	Example:
device_count	Number of devices seen in our panel during the date range whose home is in this census_block_group. Home is defined as the common nighttime location for the device over a 6-week period where nighttime is 6 pm – 7 am.	100
distance_traveled_from_home	Median distance (in meters) traveled from the geohash-7 of the home by the devices included in the device_count during the time period (**excluding any distances of 0**). We first find the median for each device and then find the median across all of the devices	200
bucketed_distance_traveled	Key is range of meters (from geohash-7 of home) and value is device count. If a device made multiple trips, we use the median distance for the device.	{"0": 100, "1-1000": 40, …"<50000": 0}

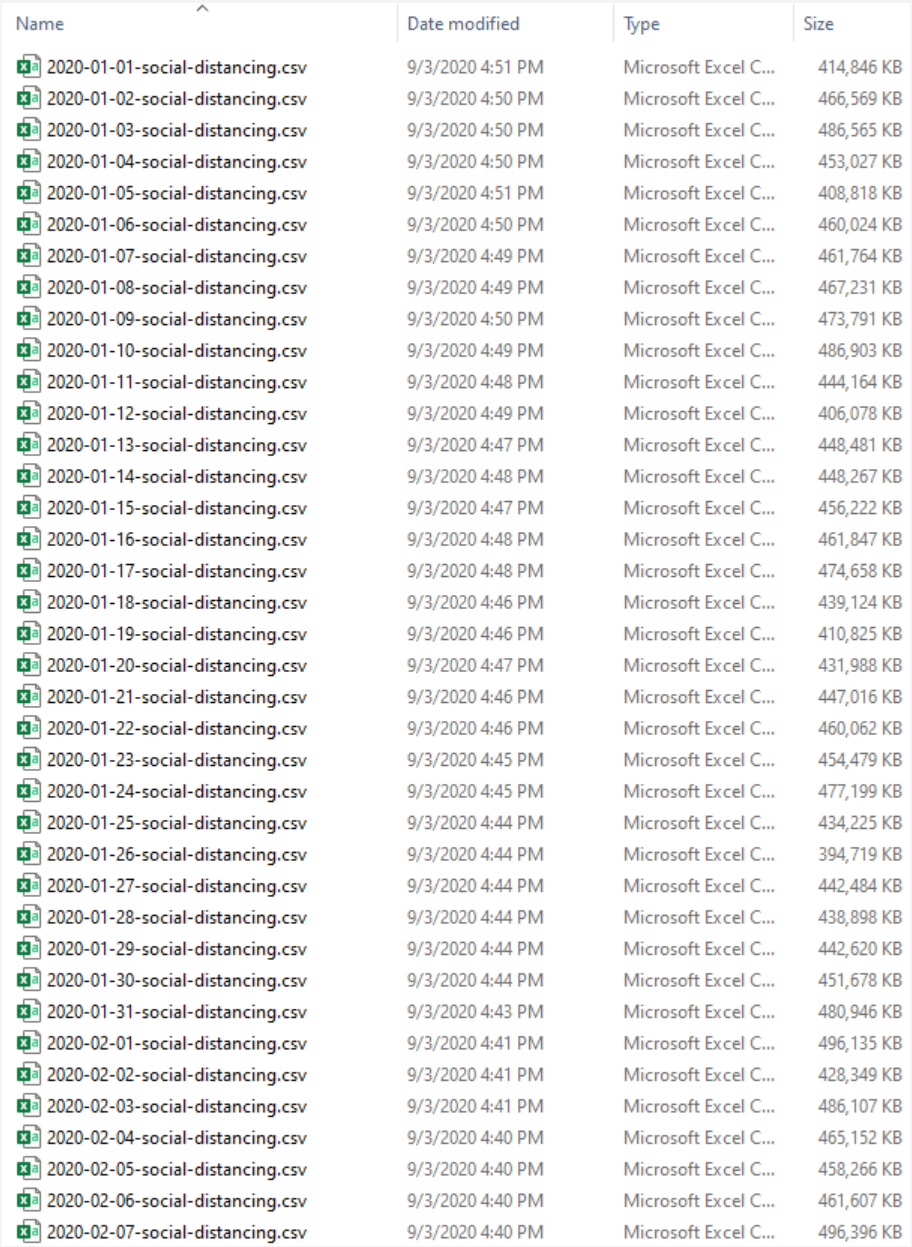
What information do Social Distancing Metrics contain?

Column Name:	Descriptions:	Example:
completely_home_device_count	Out of the device_count, the number of devices which did not leave the geohash-7 in which their home is located during the time period.	40
median_home_dwell_time	Median dwell time at home geohash-7 ("home") in minutes for all devices in the device_count during the time period. For each device, we summed the observed minutes at home across the day (whether or not these were contiguous) to get the total minutes for each device. Then we calculate the median of all these devices.	1200
bucketed_home_dwell_time	Key is range of minutes and value is device count of devices that dwelled at geohash-7 of home for the given time period. For each device, we summed the observed minutes at home across the day to get the total minutes for each device this day. Then we count how many devices are in each bucket.	{"<60": 0, "61-360": 0, "361-720": 10, "721-1080": 40, ">1081": 50}

What information do Social Distancing Metrics contain?

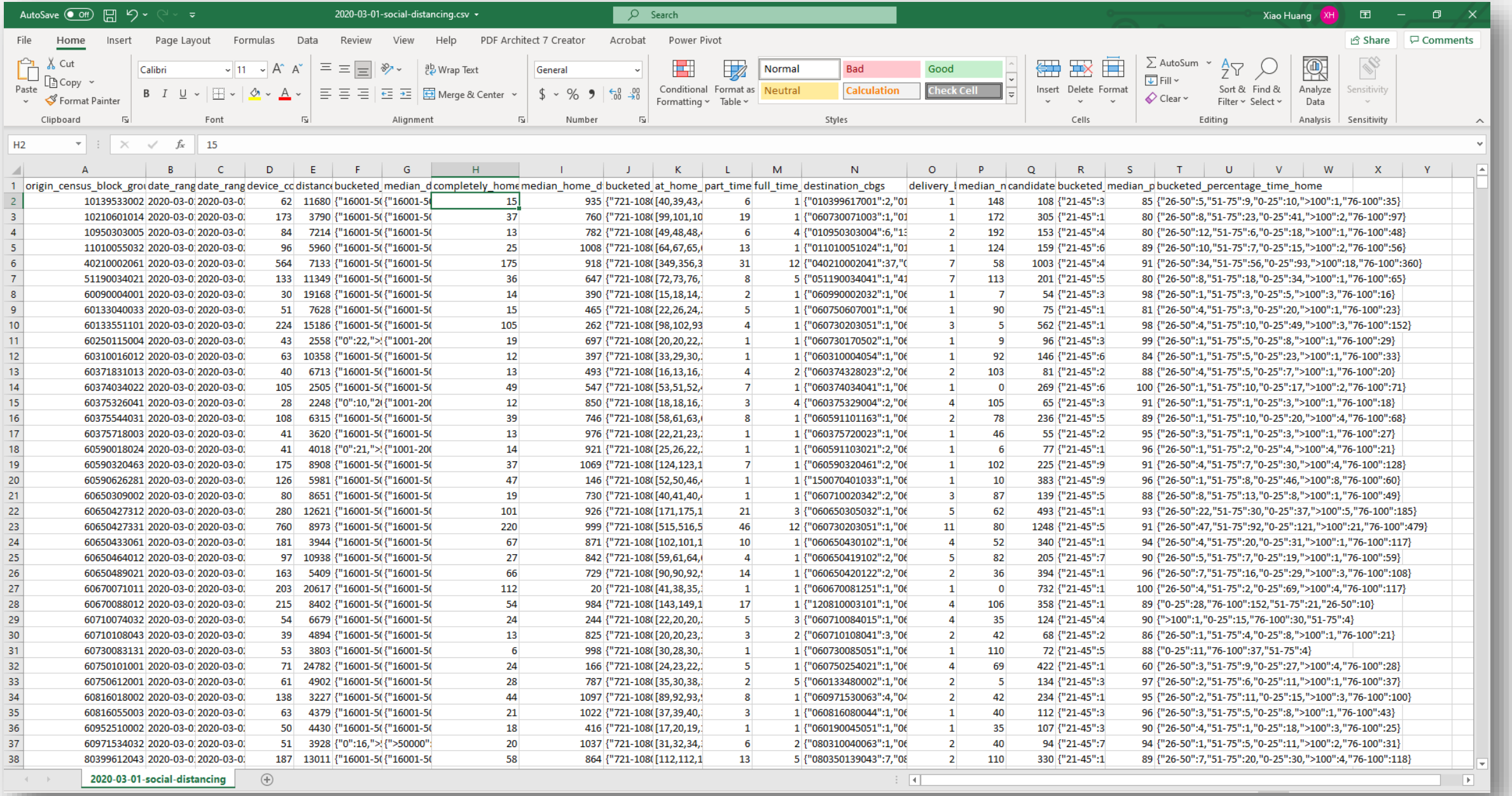
Column Name:	Descriptions:	Example:
at_home_by_each_hour	A mapping of hour of day to the number of devices at geohash-7 home in each hour over the course of the day in local time. First element in the array corresponds to the hour of midnight to 1am.	[90, 90, 90, 80, 80, 70, 70, ...]
destination_cbgs	Key is a destination census block group and value is the number of devices with a home in census_block_group that stopped in the given destination census block group for >1 minute during the time period.	{"130890212162":91, "131210101101":22, "131350502123":20}
bucketed_away_from_home_time	Key is range of minutes and value is device count of devices that dwelled anywhere outside of the geohash-7 of home for the given time period.	"0- 20": 5, "21-45": 4, "46-60": 5, "61-120": 4, "121-180": 5, "181-240": 10}

What do Social Distancing Metrics look like?



Name	Date modified	Type	Size
2020-01-01-social-distancing.csv	9/3/2020 4:51 PM	Microsoft Excel C...	414,846 KB
2020-01-02-social-distancing.csv	9/3/2020 4:50 PM	Microsoft Excel C...	466,569 KB
2020-01-03-social-distancing.csv	9/3/2020 4:50 PM	Microsoft Excel C...	486,565 KB
2020-01-04-social-distancing.csv	9/3/2020 4:50 PM	Microsoft Excel C...	453,027 KB
2020-01-05-social-distancing.csv	9/3/2020 4:51 PM	Microsoft Excel C...	408,818 KB
2020-01-06-social-distancing.csv	9/3/2020 4:50 PM	Microsoft Excel C...	460,024 KB
2020-01-07-social-distancing.csv	9/3/2020 4:49 PM	Microsoft Excel C...	461,764 KB
2020-01-08-social-distancing.csv	9/3/2020 4:49 PM	Microsoft Excel C...	467,231 KB
2020-01-09-social-distancing.csv	9/3/2020 4:50 PM	Microsoft Excel C...	473,791 KB
2020-01-10-social-distancing.csv	9/3/2020 4:49 PM	Microsoft Excel C...	486,903 KB
2020-01-11-social-distancing.csv	9/3/2020 4:48 PM	Microsoft Excel C...	444,164 KB
2020-01-12-social-distancing.csv	9/3/2020 4:49 PM	Microsoft Excel C...	406,078 KB
2020-01-13-social-distancing.csv	9/3/2020 4:47 PM	Microsoft Excel C...	448,481 KB
2020-01-14-social-distancing.csv	9/3/2020 4:48 PM	Microsoft Excel C...	448,267 KB
2020-01-15-social-distancing.csv	9/3/2020 4:47 PM	Microsoft Excel C...	456,222 KB
2020-01-16-social-distancing.csv	9/3/2020 4:48 PM	Microsoft Excel C...	461,847 KB
2020-01-17-social-distancing.csv	9/3/2020 4:48 PM	Microsoft Excel C...	474,658 KB
2020-01-18-social-distancing.csv	9/3/2020 4:46 PM	Microsoft Excel C...	439,124 KB
2020-01-19-social-distancing.csv	9/3/2020 4:46 PM	Microsoft Excel C...	410,825 KB
2020-01-20-social-distancing.csv	9/3/2020 4:47 PM	Microsoft Excel C...	431,988 KB
2020-01-21-social-distancing.csv	9/3/2020 4:46 PM	Microsoft Excel C...	447,016 KB
2020-01-22-social-distancing.csv	9/3/2020 4:46 PM	Microsoft Excel C...	460,062 KB
2020-01-23-social-distancing.csv	9/3/2020 4:45 PM	Microsoft Excel C...	454,479 KB
2020-01-24-social-distancing.csv	9/3/2020 4:45 PM	Microsoft Excel C...	477,199 KB
2020-01-25-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	434,225 KB
2020-01-26-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	394,719 KB
2020-01-27-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	442,484 KB
2020-01-28-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	438,898 KB
2020-01-29-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	442,620 KB
2020-01-30-social-distancing.csv	9/3/2020 4:44 PM	Microsoft Excel C...	451,678 KB
2020-01-31-social-distancing.csv	9/3/2020 4:43 PM	Microsoft Excel C...	480,946 KB
2020-02-01-social-distancing.csv	9/3/2020 4:41 PM	Microsoft Excel C...	496,135 KB
2020-02-02-social-distancing.csv	9/3/2020 4:41 PM	Microsoft Excel C...	428,349 KB
2020-02-03-social-distancing.csv	9/3/2020 4:41 PM	Microsoft Excel C...	486,107 KB
2020-02-04-social-distancing.csv	9/3/2020 4:40 PM	Microsoft Excel C...	465,152 KB
2020-02-05-social-distancing.csv	9/3/2020 4:40 PM	Microsoft Excel C...	458,266 KB
2020-02-06-social-distancing.csv	9/3/2020 4:40 PM	Microsoft Excel C...	461,607 KB
2020-02-07-social-distancing.csv	9/3/2020 4:40 PM	Microsoft Excel C...	496,396 KB

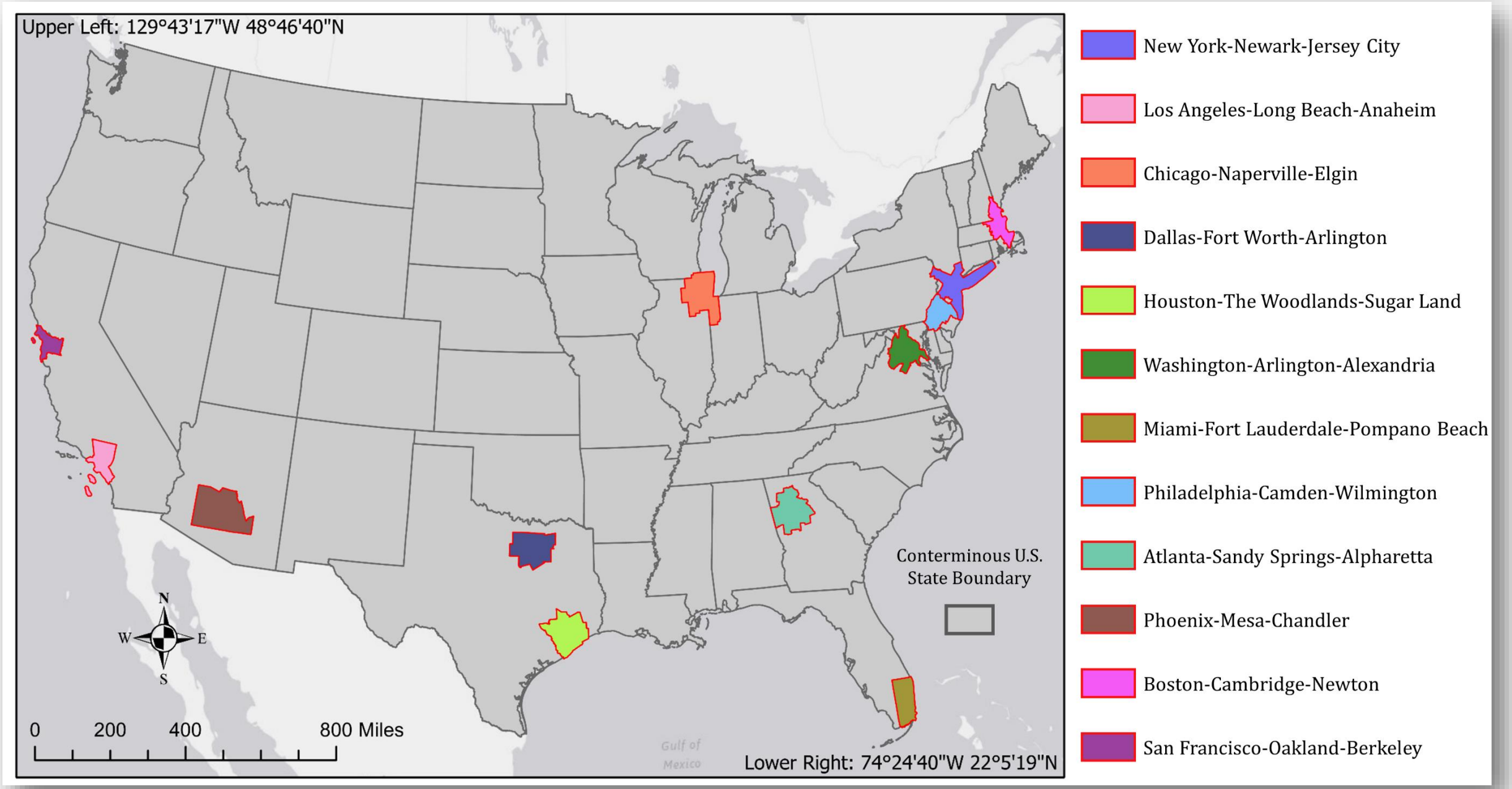
What do Social Distancing Metrics look like?

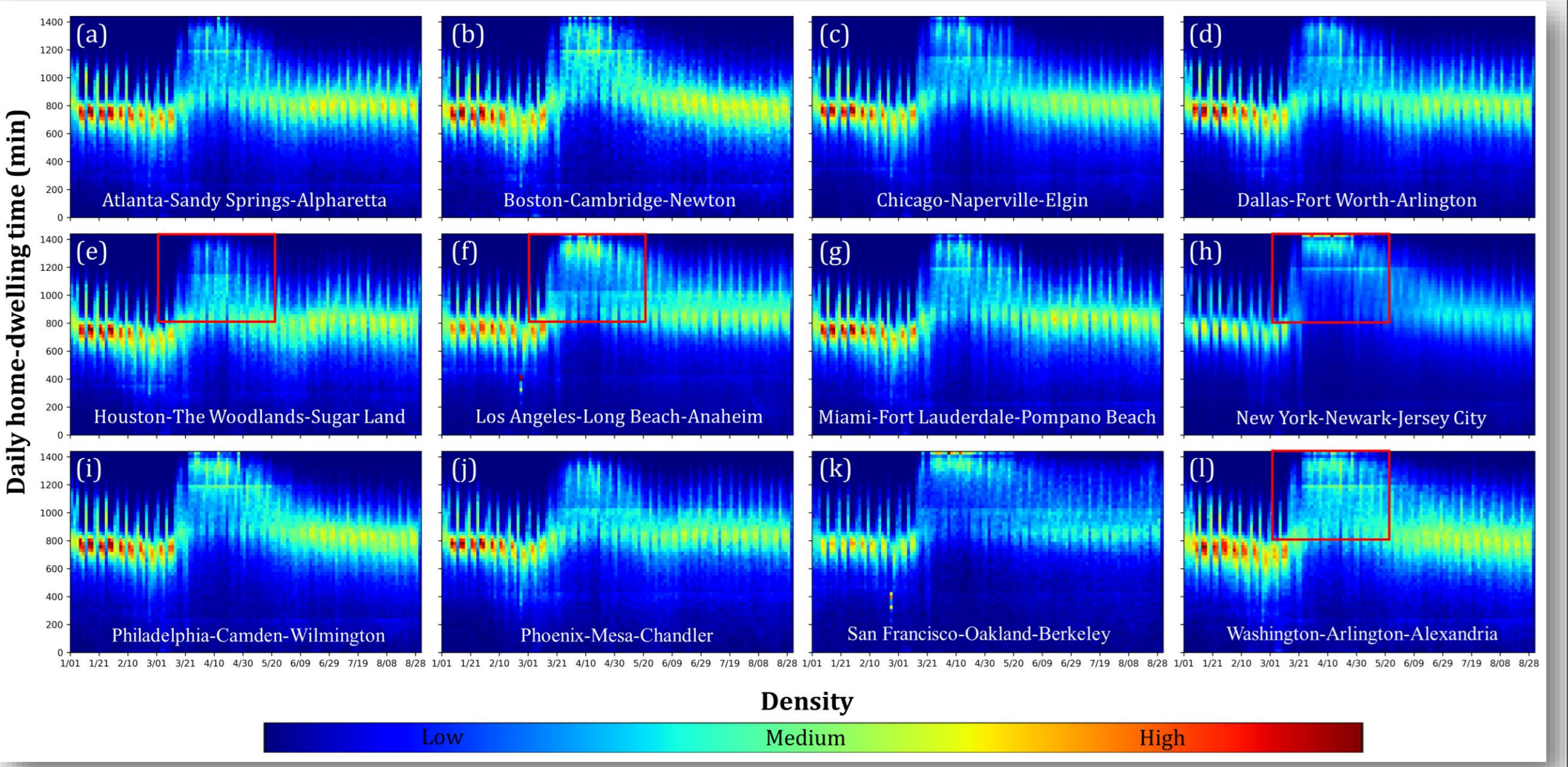


origin_census_block_gro	date_rang	date_rang	device_co	distance	bucketed_	median_d	completely_home	median_home_d	bucketed_	at_home_	part_time	full_time_	destination_cbgs	delivery_l	median_n	candidate	bucketed_	median_p	bucketed_percentage_time_home
10139533002	2020-03-0	2020-03-0	62	11680	{"16001-5(	{"16001-5(	15	935	{"721-108(	[40,39,43,4	6	1	{"010399617001":2,"01	1	148	108	{"21-45":3	85	{"26-50":5,"51-75":9,"0-25":10,">100":1,"76-100":35}
10210601014	2020-03-0	2020-03-0	173	3790	{"16001-5(	{"16001-5(	37	760	{"721-108(	[99,101,10	19	1	{"060730071003":1,"01	1	172	305	{"21-45":1	80	{"26-50":8,"51-75":23,"0-25":41,">100":2,"76-100":97}
10950303005	2020-03-0	2020-03-0	84	7214	{"16001-5(	{"16001-5(	13	782	{"721-108(	[49,48,48,4	6	4	{"010950303004":6,"13	2	192	153	{"21-45":4	80	{"26-50":12,"51-75":6,"0-25":18,">100":1,"76-100":48}
11010055032	2020-03-0	2020-03-0	96	5960	{"16001-5(	{"16001-5(	25	1008	{"721-108(	[64,67,65,(	13	1	{"011010051024":1,"01	1	124	159	{"21-45":6	89	{"26-50":10,"51-75":7,"0-25":15,">100":2,"76-100":56}
40210002061	2020-03-0	2020-03-0	564	7133	{"16001-5(	{"16001-5(	175	918	{"721-108(	[349,356,3	31	12	{"040210002041":37,"(	7	58	1003	{"21-45":4	91	{"26-50":34,"51-75":56,"0-25":93,">100":18,"76-100":360}
51190034021	2020-03-0	2020-03-0	133	11349	{"16001-5(	{"16001-5(	36	647	{"721-108(	[72,73,76,	8	5	{"051190034041":1,"41	7	113	201	{"21-45":5	80	{"26-50":8,"51-75":18,"0-25":34,">100":1,"76-100":65}
60090004001	2020-03-0	2020-03-0	30	19168	{"16001-5(	{"16001-5(	14	390	{"721-108(	[15,18,14,	2	1	{"060990002032":1,"06	1	7	54	{"21-45":3	98	{"26-50":1,"51-75":3,"0-25":5,">100":3,"76-100":16}
60133040033	2020-03-0	2020-03-0	51	7628	{"16001-5(	{"16001-5(	15	465	{"721-108(	[22,26,24,	5	1	{"060750607001":1,"06	1	90	75	{"21-45":1	81	{"26-50":4,"51-75":3,"0-25":20,">100":1,"76-100":23}
60133551101	2020-03-0	2020-03-0	224	15186	{"16001-5(	{"16001-5(	105	262	{"721-108(	[98,102,93	4	1	{"060730203051":1,"06	3	5	562	{"21-45":1	98	{"26-50":4,"51-75":10,"0-25":49,">100":3,"76-100":152}
60250115004	2020-03-0	2020-03-0	43	2558	{"0":22,">	{"1001-20(	19	697	{"721-108(	[20,20,22,	1	1	{"060730170502":1,"06	1	9	96	{"21-45":3	99	{"26-50":1,"51-75":5,"0-25":8,">100":1,"76-100":29}
60310016012	2020-03-0	2020-03-0	63	10358	{"16001-5(	{"16001-5(	12	397	{"721-108(	[33,29,30,	1	1	{"060310004054":1,"06	1	92	146	{"21-45":6	84	{"26-50":1,"51-75":5,"0-25":23,">100":1,"76-100":33}
60371831013	2020-03-0	2020-03-0	40	6713	{"16001-5(	{"16001-5(	13	493	{"721-108(	[16,13,16,	4	2	{"060374328023":2,"06	2	103	81	{"21-45":2	88	{"26-50":4,"51-75":5,"0-25":7,">100":1,"76-100":20}
60374034022	2020-03-0	2020-03-0	105	2505	{"16001-5(	{"16001-5(	49	547	{"721-108(	[53,51,52,4	7	1	{"060374034041":1,"06	1	0	269	{"21-45":6	100	{"26-50":1,"51-75":10,"0-25":17,">100":2,"76-100":71}
60375326041	2020-03-0	2020-03-0	28	2248	{"0":10,"2(	{"1001-20(	12	850	{"721-108(	[18,18,16,	3	4	{"060375329004":2,"06	4	105	65	{"21-45":3	91	{"26-50":1,"51-75":1,"0-25":3,">100":1,"76-100":18}
60375544031	2020-03-0	2020-03-0	108	6315	{"16001-5(	{"16001-5(	39	746	{"721-108(	[58,61,63,4	8	1	{"060591101163":1,"06	2	78	236	{"21-45":5	89	{"26-50":1,"51-75":10,"0-25":20,">100":4,"76-100":68}
60375718003	2020-03-0	2020-03-0	41	3620	{"16001-5(	{"16001-5(	13	976	{"721-108(	[22,21,23,	1	1	{"060375720023":1,"06	1	46	55	{"21-45":2	95	{"26-50":3,"51-75":1,"0-25":3,">100":1,"76-100":27}
60590018024	2020-03-0	2020-03-0	41	4018	{"0":21,">	{"1001-20(	14	921	{"721-108(	[25,26,22,	1	1	{"060591103021":2,"06	1	6	77	{"21-45":1	96	{"26-50":1,"51-75":2,"0-25":4,">100":4,"76-100":21}
60590320463	2020-03-0	2020-03-0	175	8908	{"16001-5(	{"16001-5(	37	1069	{"721-108(	[124,123,1	7	1	{"060590320461":2,"06	1	102	225	{"21-45":9	91	{"26-50":4,"51-75":7,"0-25":30,">100":4,"76-100":128}
60590626281	2020-03-0	2020-03-0	126	5981	{"16001-5(	{"16001-5(	47	146	{"721-108(	[52,50,46,4	1	1	{"150070401033":1,"06	1	10	383	{"21-45":9	96	{"26-50":1,"51-75":8,"0-25":46,">100":8,"76-100":60}
60650309002	2020-03-0	2020-03-0	80	8651	{"16001-5(	{"16001-5(	19	730	{"721-108(	[40,41,40,4	1	1	{"060710020342":2,"06	3	87	139	{"21-45":5	88	{"26-50":8,"51-75":13,"0-25":8,">100":1,"76-100":49}
60650427312	2020-03-0	2020-03-0	280	12621	{"16001-5(	{"16001-5(	101	926	{"721-108(	[171,175,1	21	3	{"060650305032":1,"06	5	62	493	{"21-45":1	93	{"26-50":22,"51-75":30,"0-25":37,">100":5,"76-100":185}
60650427331	2020-03-0	2020-03-0	760	8973	{"16001-5(	{"16001-5(	220	999	{"721-108(	[515,516,5	46	12	{"060730203051":1,"06	11	80	1248	{"21-45":5	91	{"26-50":47,"51-75":92,"0-25":121,">100":21,"76-100":479}
60650433061	2020-03-0	2020-03-0	181	3944	{"16001-5(	{"16001-5(	67	871	{"721-108(	[102,101,1	10	1	{"060650430102":1,"06	4	52	340	{"21-45":1	94	{"26-50":4,"51-75":20,"0-25":31,">100":1,"76-100":117}
60650464012	2020-03-0	2020-03-0	97	10938	{"16001-5(	{"16001-5(	27	842	{"721-108(	[59,61,64,4	4	1	{"060650419102":2,"06	5	82	205	{"21-45":7	90	{"26-50":5,"51-75":7,"0-25":19,">100":1,"76-100":59}
60650489021	2020-03-0	2020-03-0	163	5409	{"16001-5(	{"16001-5(	66	729	{"721-108(	[90,90,92,9	14	1	{"060650420122":2,"06	2	36	394	{"21-45":1	96	{"26-50":7,"51-75":16,"0-25":29,">100":3,"76-100":108}
60670071011	2020-03-0	2020-03-0	203	20617	{"16001-5(	{"16001-5(	112	20	{"721-108(	[41,38,35,	1	1	{"060670081251":1,"06	1	0	732	{"21-45":1	100	{"26-50":4,"51-75":2,"0-25":69,">100":4,"76-100":117}
60670088012	2020-03-0	2020-03-0	215	8402	{"16001-5(	{"16001-5(	54	984	{"721-108(	[143,149,1	17	1	{"120810003101":1,"06	4	106	358	{"21-45":1	89	{"0-25":28,"76-100":152,"51-75":21,"26-50":10}
60710074032	2020-03-0	2020-03-0	54	6679	{"16001-5(	{"16001-5(	24	244	{"721-108(	[22,20,20,	5	3	{"060710084015":1,"06	4	35	124	{"21-45":4	90	{">100":1,"0-25":15,"76-100":30,"51-75":4}
60710108043	2020-03-0	2020-03-0	39	4894	{"16001-5(	{"16001-5(	13	825	{"721-108(	[20,20,23,	3	2	{"060710108041":3,"06	2	42	68	{"21-45":2	86	{"26-50":1,"51-75":4,"0-25":8,">100":1,"76-100":21}
60730083131	2020-03-0	2020-03-0	53	3803	{"16001-5(	{"16001-5(	6	998	{"721-108(	[30,28,30,	1	1	{"060730085051":1,"06	1	110	72	{"21-45":5	88	{"0-25":11,"76-100":37,"51-75":4}
60750101001	2020-03-0	2020-03-0	71	24782	{"16001-5(	{"16001-5(	24	166	{"721-108(	[24,23,22,	5	1	{"060750254021":1,"06	4	69	422	{"21-45":1	60	{"26-50":3,"51-75":9,"0-25":27,">100":4,"76-100":28}
60750612001	2020-03-0	2020-03-0	61	4902	{"16001-5(	{"16001-5(	28	787	{"721-108(	[35,30,38,	2	5	{"060133480002":1,"06	2	5	134	{"21-45":3	97	{"26-50":2,"51-75":6,"0-25":11,">100":1,"76-100":37}
60816018002	2020-03-0	2020-03-0	138	3227	{"16001-5(	{"16001-5(	44	1097	{"721-108(	[89,92,93,9	8	1	{"060971530063":4,"04	2	42	234	{"21-45":1	95	{"26-50":2,"51-75":11,"0-25":15,">100":3,"76-100":100}
60816055003	2020-03-0	2020-03-0	63	4379	{"16001-5(	{"16001-5(	21	1022	{"721-108(	[37,39,40,	3	1	{"060816080044":1,"06	1	40	112	{"21-45":3	96	{"26-50":3,"51-75":5,"0-25":8,">100":1,"76-100":43}
60952510002	2020-03-0	2020-03-0	50	4430	{"16001-5(	{"16001-5(	18	416	{"721-108(	[17,20,19,	1	1	{"060190045051":1,"06	1	35	107	{"21-45":3	90	{"26-50":4,"51-75":1,"0-25":18,">100":3,"76-100":25}
60971534032	2020-03-0	2020-03-0	51	3928	{"0":16,">	{">50000"	20	1037	{"721-108(	[31,32,34,	6	2	{"080310040063":1,"06	2	40	94	{"21-45":7	94	{"26-50":1,"51-75":5,"0-25":11,">100":2,"76-100":31}
80399612043	2020-03-0	2020-03-0	187	13011	{"16001-5(	{"16001-5(	58	864	{"721-108(	[112,112,1	13	5	{"080350139043":7,"08	2	110	330	{"21-45":1	89	{"26-50":7,"51-75":20,"0-25":30,">100":4,"76-100":118}

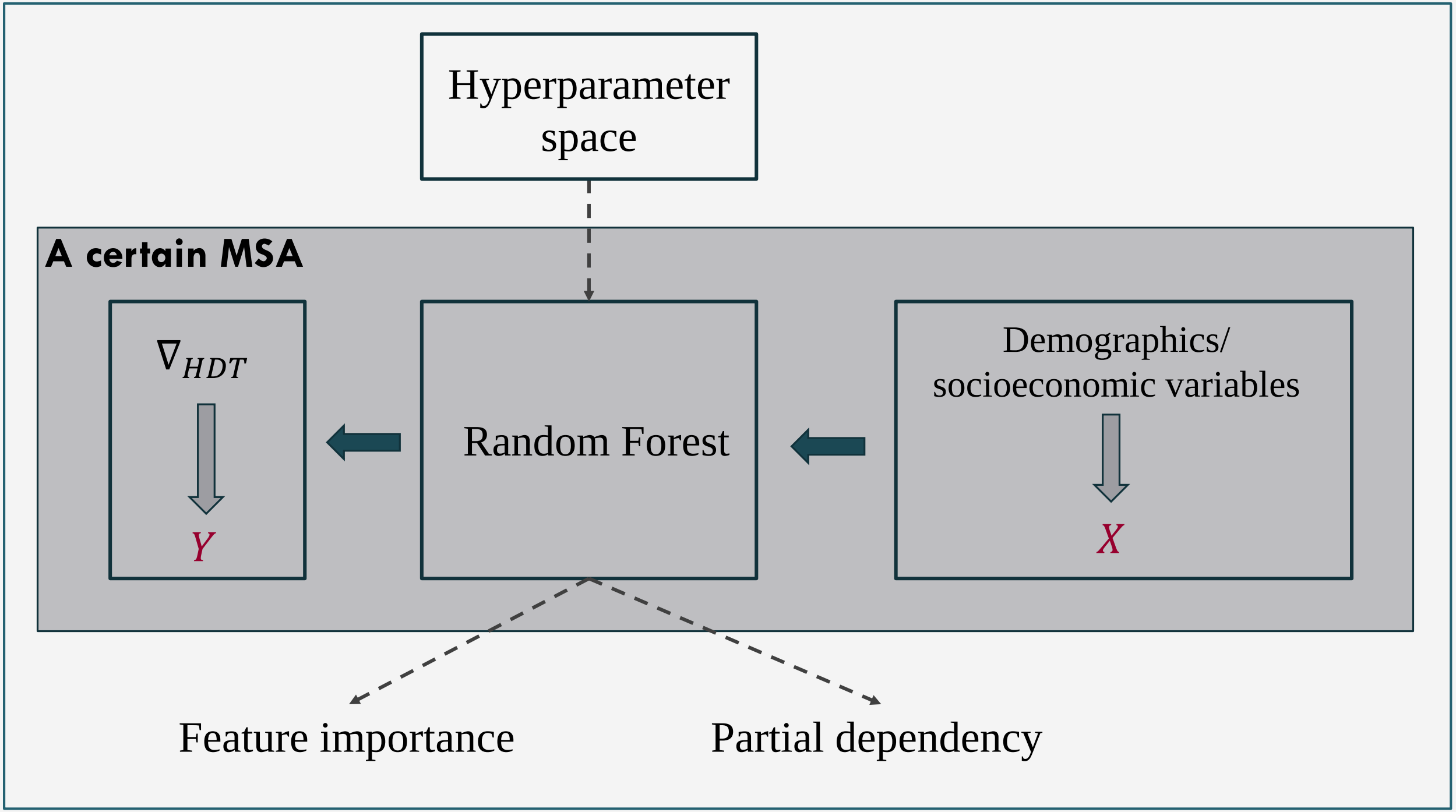
Staying at home is a privilege: evidence from fine-grained mobile phone location data in the U.S. during the COVID-19 pandemic

- We use **fine-grained home-dwelling records** collected from millions of mobile devices to assess and cross-compare the **compliance** of stay-at-home orders in the **top twelve most-populated MSAs** in the U.S.
- We apply an **optimized Random Forest** algorithm, a popular machine learning method, to statistically investigate the **contribution of demographic/socioeconomic variables** to the **increase** in home-dwelling time.
- We present the **feature importance** of selected demographic/socioeconomic variables and the performance of the designed Random Forest model in predicting the increase in home-dwelling time based on these variables.


Upper Left: 129°43'17"W 48°46'40"N
Lower Right: 74°24'40"W 22°5'19"N
N
W
E
S
0
200
400
800 Miles
Gulf of Mexico
Conterminous U.S. State Boundary
New York-Newark-Jersey City
Los Angeles-Long Beach-Anaheim
Chicago-Naperville-Elgin
Dallas-Fort Worth-Arlington
Houston-The Woodlands-Sugar Land
Washington-Arlington-Alexandria
Miami-Fort Lauderdale-Pompano Beach
Philadelphia-Camden-Wilmington
Atlanta-Sandy Springs-Alpharetta
Phoenix-Mesa-Chandler
Boston-Cambridge-Newton
San Francisco-Oakland-Berkeley


Daily home-dwelling time (min)
(a)
Atlanta-Sandy Springs-Alpharetta
(b)
Boston-Cambridge-Newton
(c)
Chicago-Naperville-Elgin
(d)
Dallas-Fort Worth-Arlington
(e)
Houston-The Woodlands-Sugar Land
(f)
Los Angeles-Long Beach-Anaheim
(g)
Miami-Fort Lauderdale-Pompano Beach
(h)
New York-Newark-Jersey City
(i)
Philadelphia-Camden-Wilmington
(j)
Phoenix-Mesa-Chandler
(k)
San Francisco-Oakland-Berkeley
(l)
Washington-Arlington-Alexandria
0
200
400
600
800
1000
1200
1400
1/01 1/21 2/10 3/01 3/21 4/10 4/30 5/20 6/09 6/29 7/19 8/08 8/28
Density
Low
Medium
High

∇_{HDT}


Hyperparameter space
A certain MSA
∇_{HDT}
Y
Random Forest
Demographics/ socioeconomic variables
X
Feature importance
Partial dependency

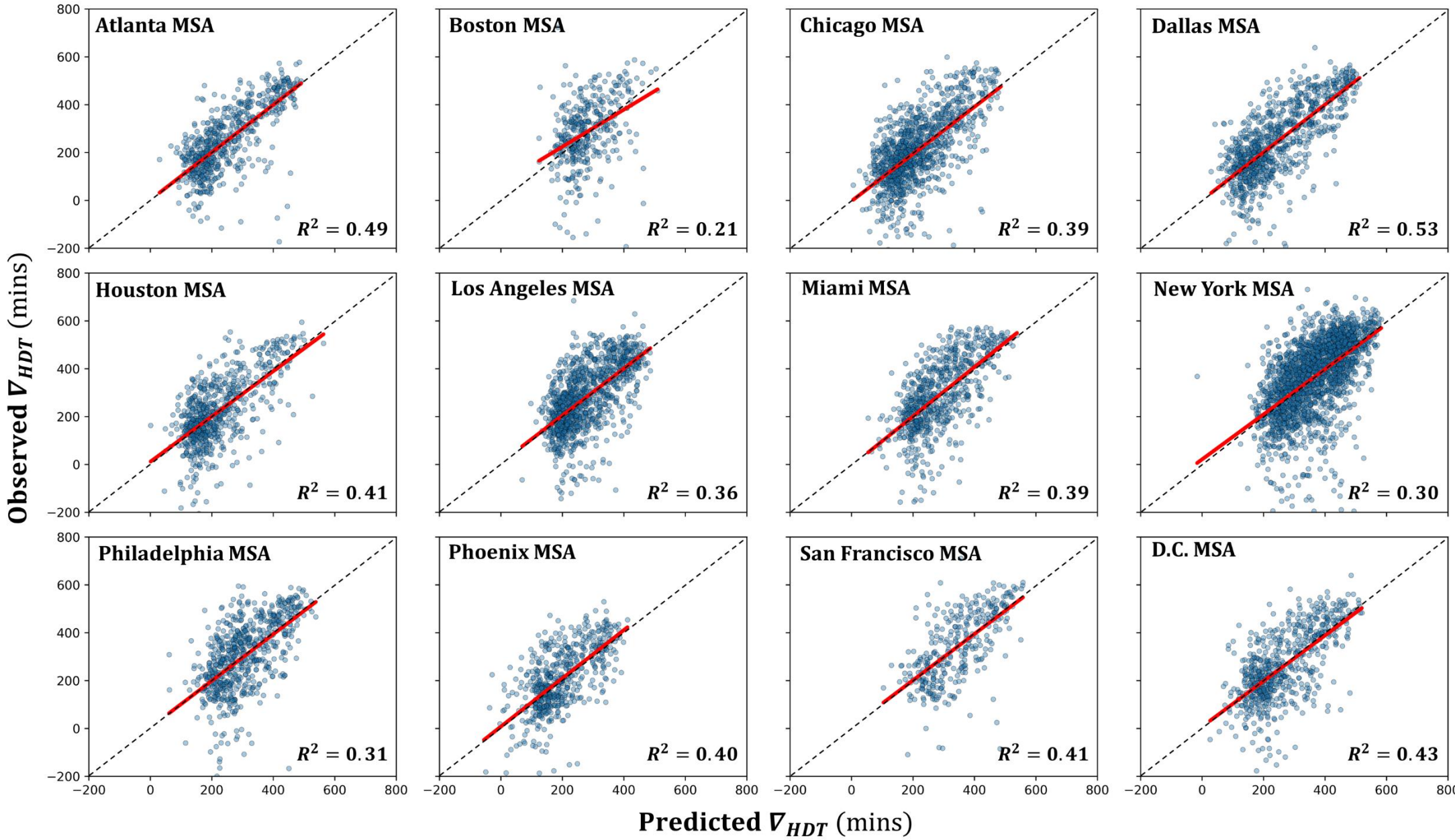
(1) A total of n_{tree} bootstrap samples, i.e., $S_1, S_2 \ldots S_{n_{tree}}$, are drawn with replacement from the training set in that MSA. A bootstrap subset contains approximately one-third of the records in the training set. The elements not included in a bootstrap subset are referred to as out-of-bag (OOB) data.

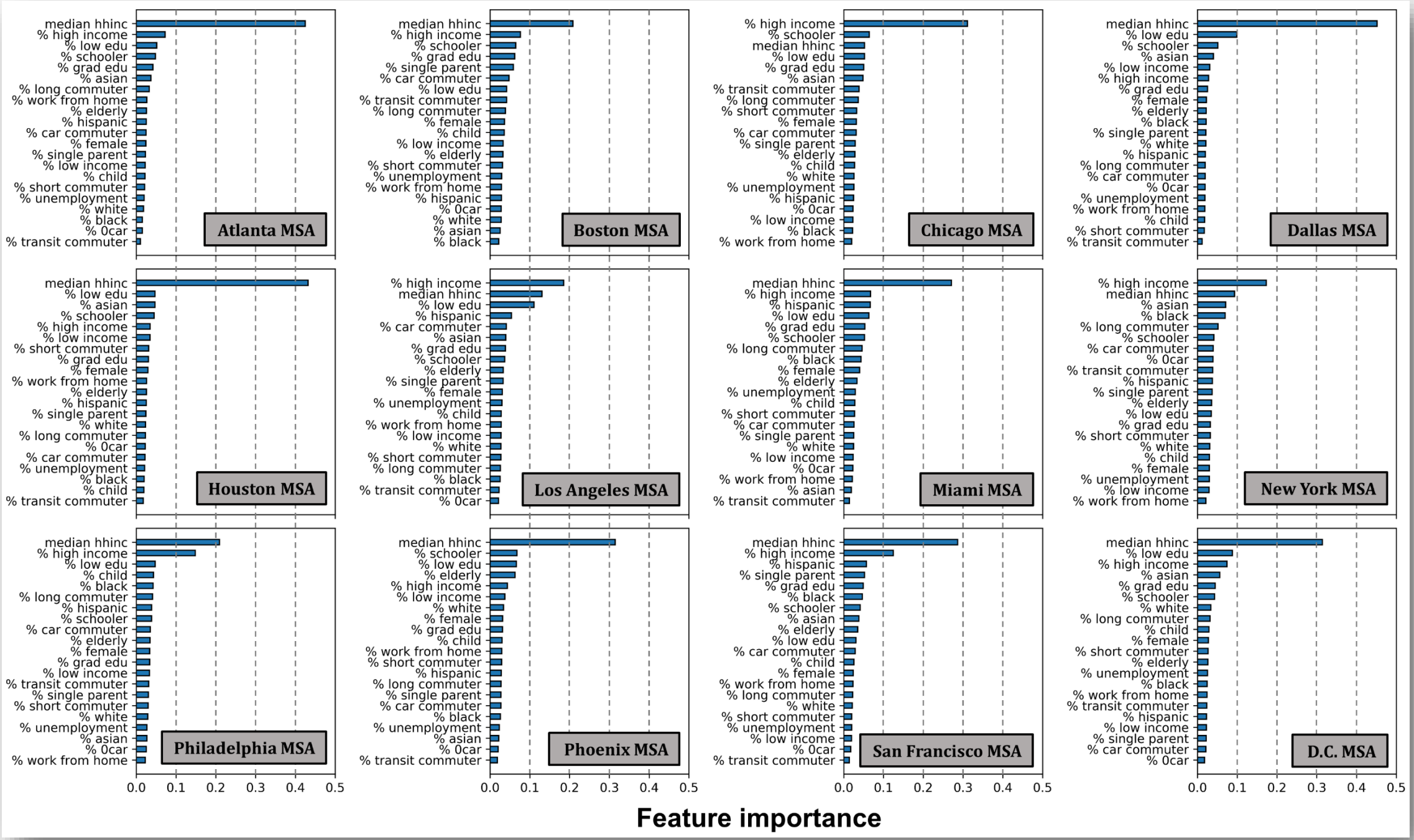
(2) The bootstrap samples are used to grow an unpruned regression tree: at each node, a total of m_{try} predictor variables are randomly selected, and the best split is chosen from among these variables. A prediction function is therefore formed for each bootstrap subset: $\widehat{Y_1} = \hat{f}(X, S_1)$, $\widehat{Y_2} = \hat{f}(X, S_2)$,… $\widehat{Y_{n_{tree}}} = \hat{f}(X, S_{n_{tree}})$.

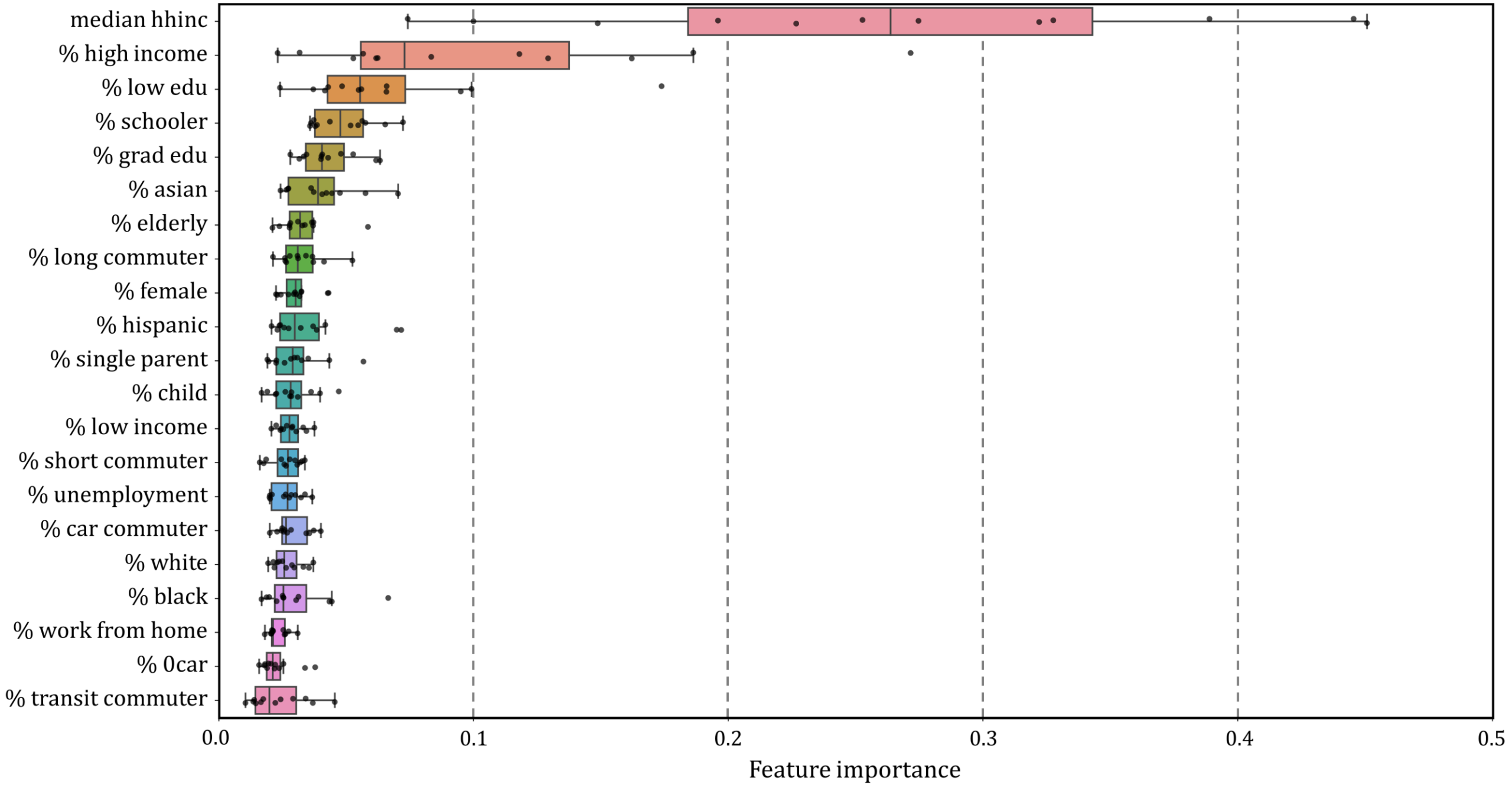
(3) The OOB data are predicted by averaging the predictions from n_{tree} trees. $\hat{Y} = \frac{1}{n_{tree}} \sum_{k=1}^{n_{tree}} \hat{f}\,(X, S_k\,)$. The importance of each predictor is further measured by calculating the percent increase in Mean Square Error (MSE).

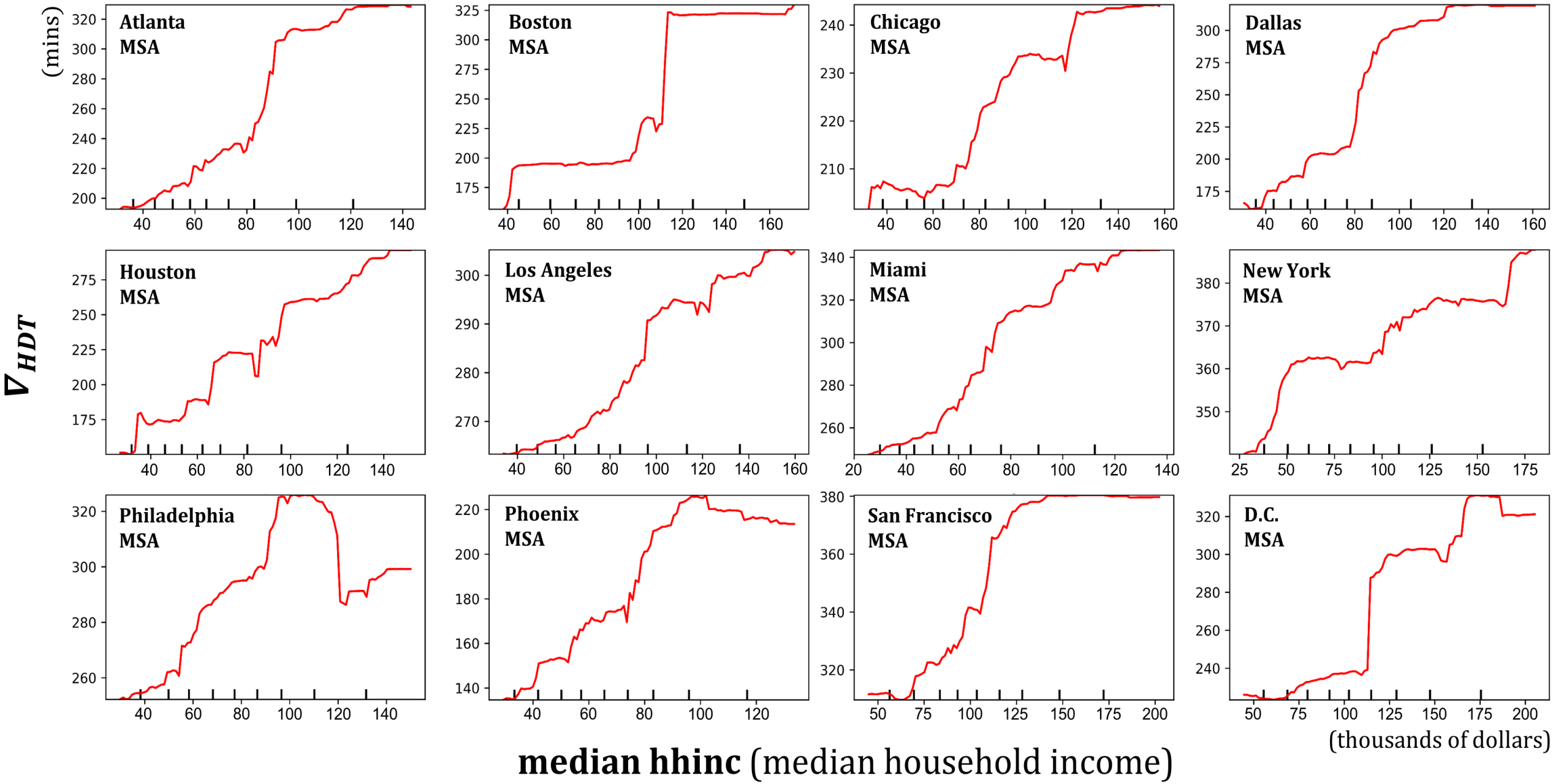
n_{tree}: [10,10,1000]

$m_{try} \in \{S,\ \frac{S}{2},\ \frac{S}{3}, \sqrt{S},\ log_2(S)\}.$


Atlanta MSA
$R^2 = 0.49$
Boston MSA
$R^2 = 0.21$
Chicago MSA
$R^2 = 0.39$
Dallas MSA
$R^2 = 0.53$
Houston MSA
$R^2 = 0.41$
Los Angeles MSA
$R^2 = 0.36$
Miami MSA
$R^2 = 0.39$
New York MSA
$R^2 = 0.30$
Philadelphia MSA
$R^2 = 0.31$
Phoenix MSA
$R^2 = 0.40$
San Francisco MSA
$R^2 = 0.41$
D.C. MSA
$R^2 = 0.43$
Observed ∇_{HDT} (mins)
Predicted ∇_{HDT} (mins)
800
600
400
200
0
−200


median hhinc
% high income
% low edu
% schooler
% grad edu
% asian
% long commuter
% work from home
% elderly
% hispanic
% car commuter
% female
% single parent
% low income
% child
% short commuter
% unemployment
% white
% black
% 0car
% transit commuter
Atlanta MSA
Boston MSA
Chicago MSA
Dallas MSA
Houston MSA
Los Angeles MSA
Miami MSA
New York MSA
Philadelphia MSA
Phoenix MSA
San Francisco MSA
D.C. MSA
0.0
0.1
0.2
0.3
0.4
0.5
Feature importance

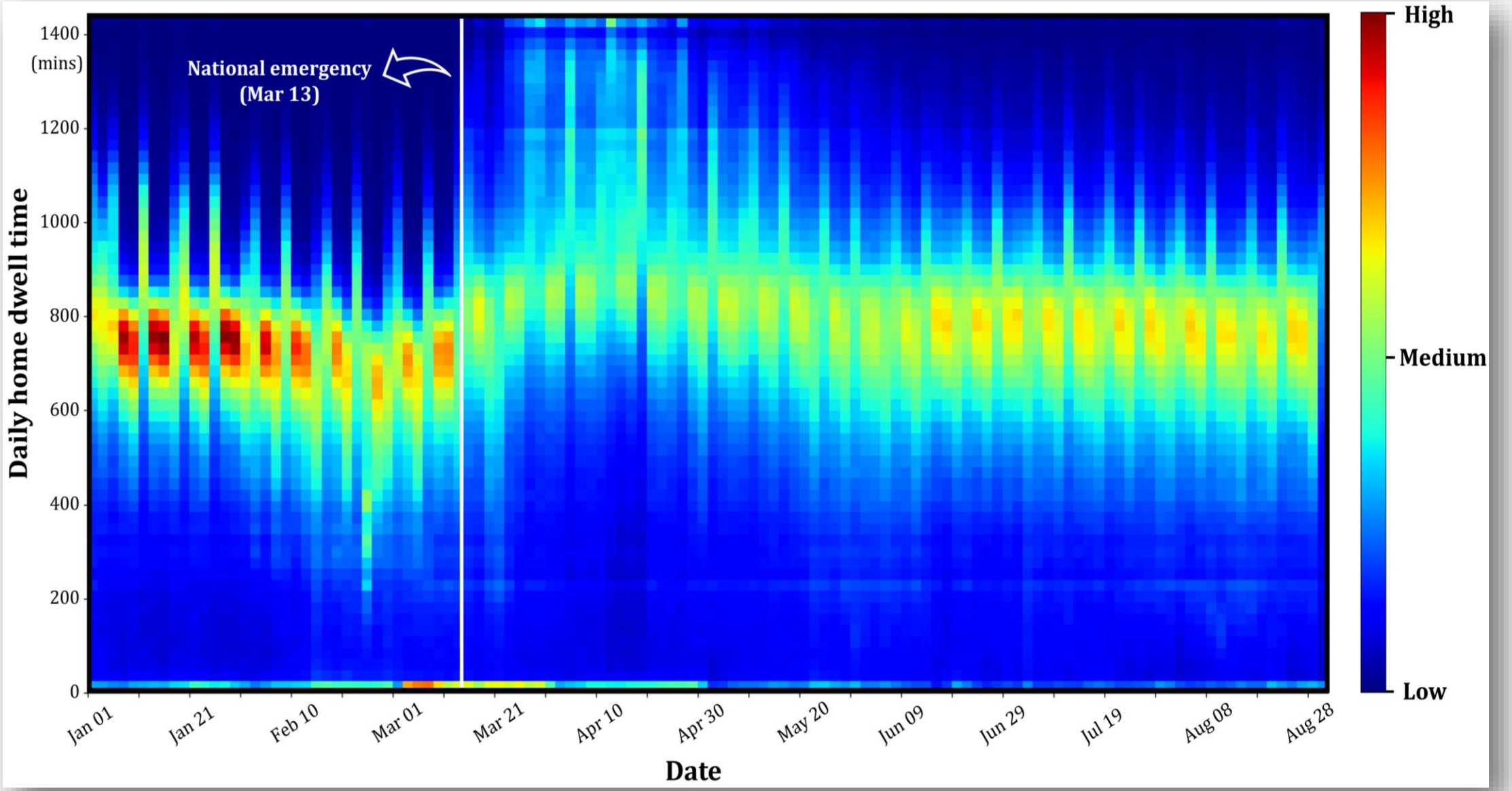

median hhinc
% high income
% low edu
% schooler
% grad edu
% asian
% elderly
% long commuter
% female
% hispanic
% single parent
% child
% low income
% short commuter
% unemployment
% car commuter
% white
% black
% work from home
% 0car
% transit commuter
0.0
0.1
0.2
0.3
0.4
0.5
Feature importance


∇_{HDT} (mins)
Atlanta MSA
Boston MSA
Chicago MSA
Dallas MSA
Houston MSA
Los Angeles MSA
Miami MSA
New York MSA
Philadelphia MSA
Phoenix MSA
San Francisco MSA
D.C. MSA
median hhinc (median household income)
(thousands of dollars)

Exploring home-dwelling time patterns during the COVID-19 pandemic via Bayesian inference: policy compliance, spatial disparity, and exposed inequity

- We document the **multi-scale spatial disparity** in social distancing compliance in the U.S, revealed by the home-dwelling time records collected from massive mobile devices.

- We **identify different social distancing stages** by performing an automatic changepoint detection on U.S. home-dwelling time records via **Bayesian inference** with weakly informative priors.



Heat map of home dwell time for **219,972 CBGs** in the U.S. from January 1, 2020 to August 31, 2020

Methodology

$\boldsymbol{\tau}$: changepoints

$\boldsymbol{\mu}$: stage means

$$\mathcal{P}(\mu_1,..,\mu_n,\tau_1,..,\tau_{n-1},\sigma|\boldsymbol{X})$$

$$= \frac{\mathcal{P}(\boldsymbol{X}|\mu_1,..,\mu_n,\tau_1,..,\tau_{n-1},\sigma)\mathcal{P}(\mu_1,..,\mu_n,\tau_1,..,\tau_{n-1},\sigma)}{\mathcal{P}(\boldsymbol{X})}$$

$$\mathcal{P}(\boldsymbol{X}|\mu_1,..,\mu_n,\tau_1,..,\tau_{n-1},\sigma)$$

$$= \prod_{i=1}^{\tau_1} \mathcal{P}(x_i|\mu_1,\sigma) \prod_{j=\tau_1+1}^{\tau_2} \mathcal{P}(x_j|\mu_2,\sigma)\ldots \prod_{m=\tau_{n-1}+1}^{N} \mathcal{P}(x_m|\mu_n,\sigma)$$

$$x_i = \begin{cases} N(\mu_1,\ \Sigma), & t \le \tau_1 \\ N(\mu_2,\ \Sigma), & \tau_1 < t < \tau_2 \\ N(\mu_3,\ \Sigma), & t \ge \tau_2 \end{cases}$$

$$\mu_1 \sim \mathcal{N}(700, \sigma)$$

$$\mu_2 \sim \mathcal{N}(800, \sigma)$$

$$\mu_2 \sim \mathcal{N}(700, \sigma)$$

$$\tau_1 \sim \text{Unif}_{dct}(50, 130)$$

$$\tau_2 \sim \text{Unif}_{dct}(\tau_1, 200)$$

τ_1: changepoint A

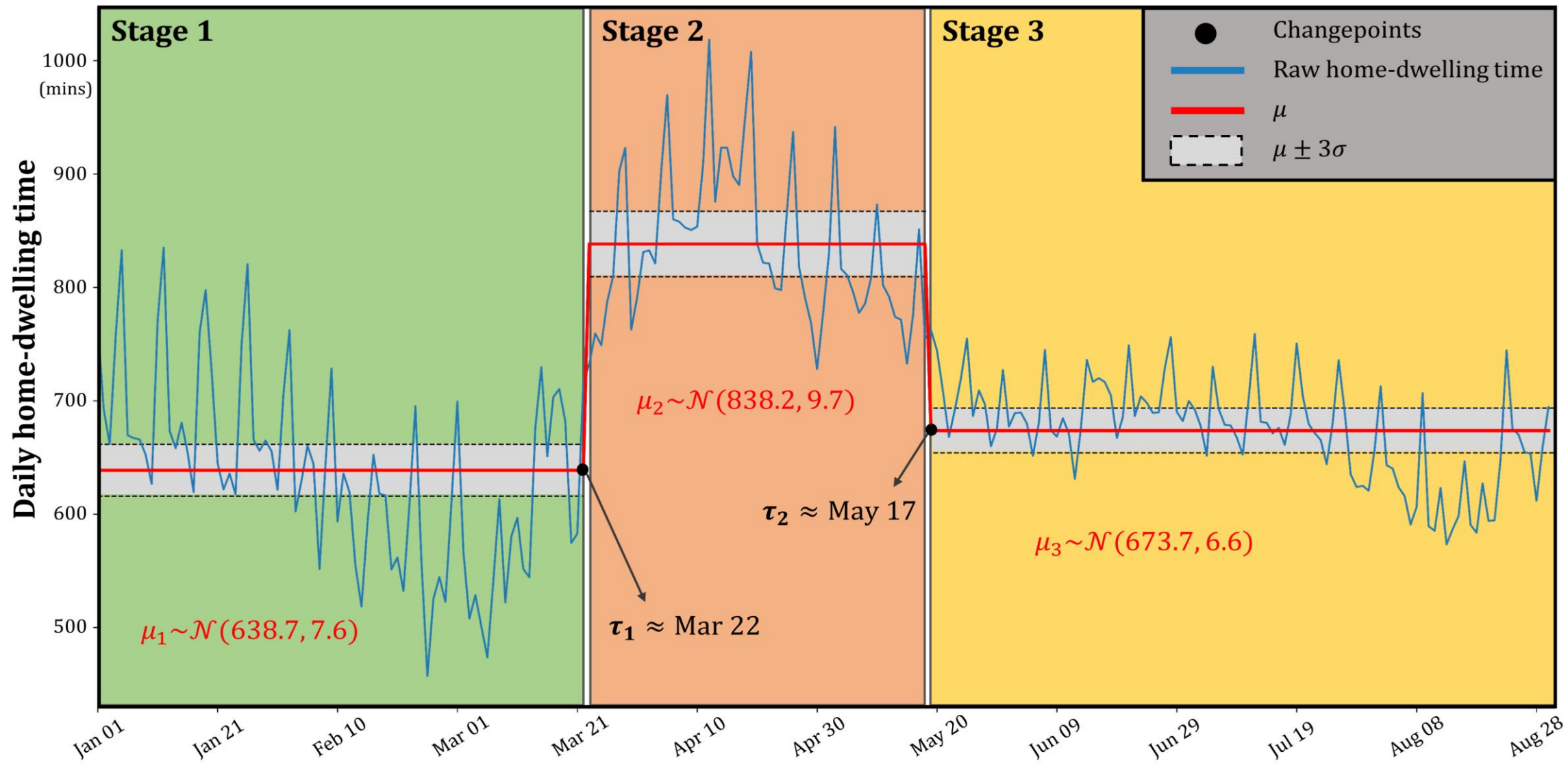
τ_2: changepoint B

μ_1: Stage 1 mean

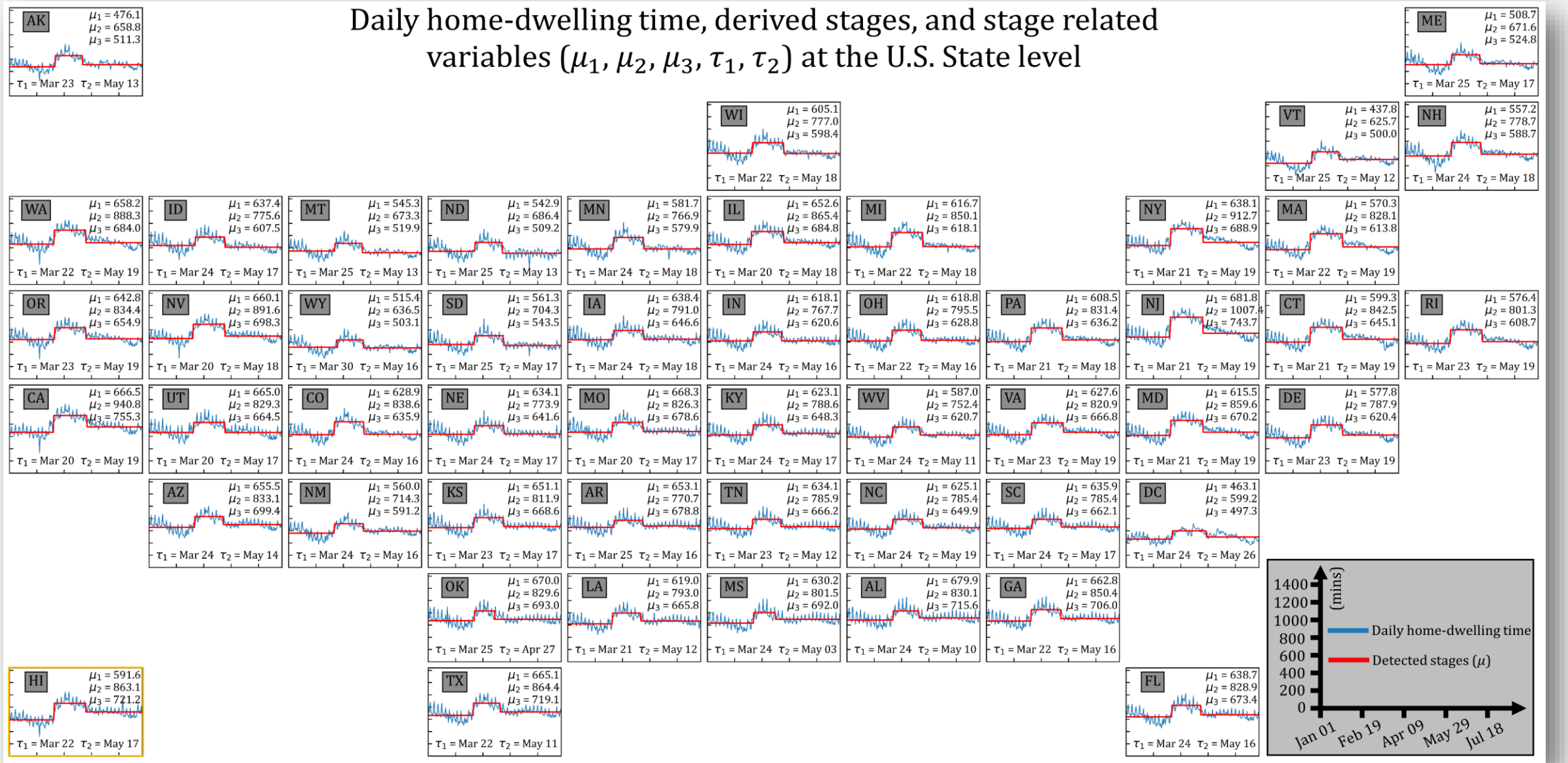
μ_2: Stage 2 mean

μ_3: Stage 3 mean

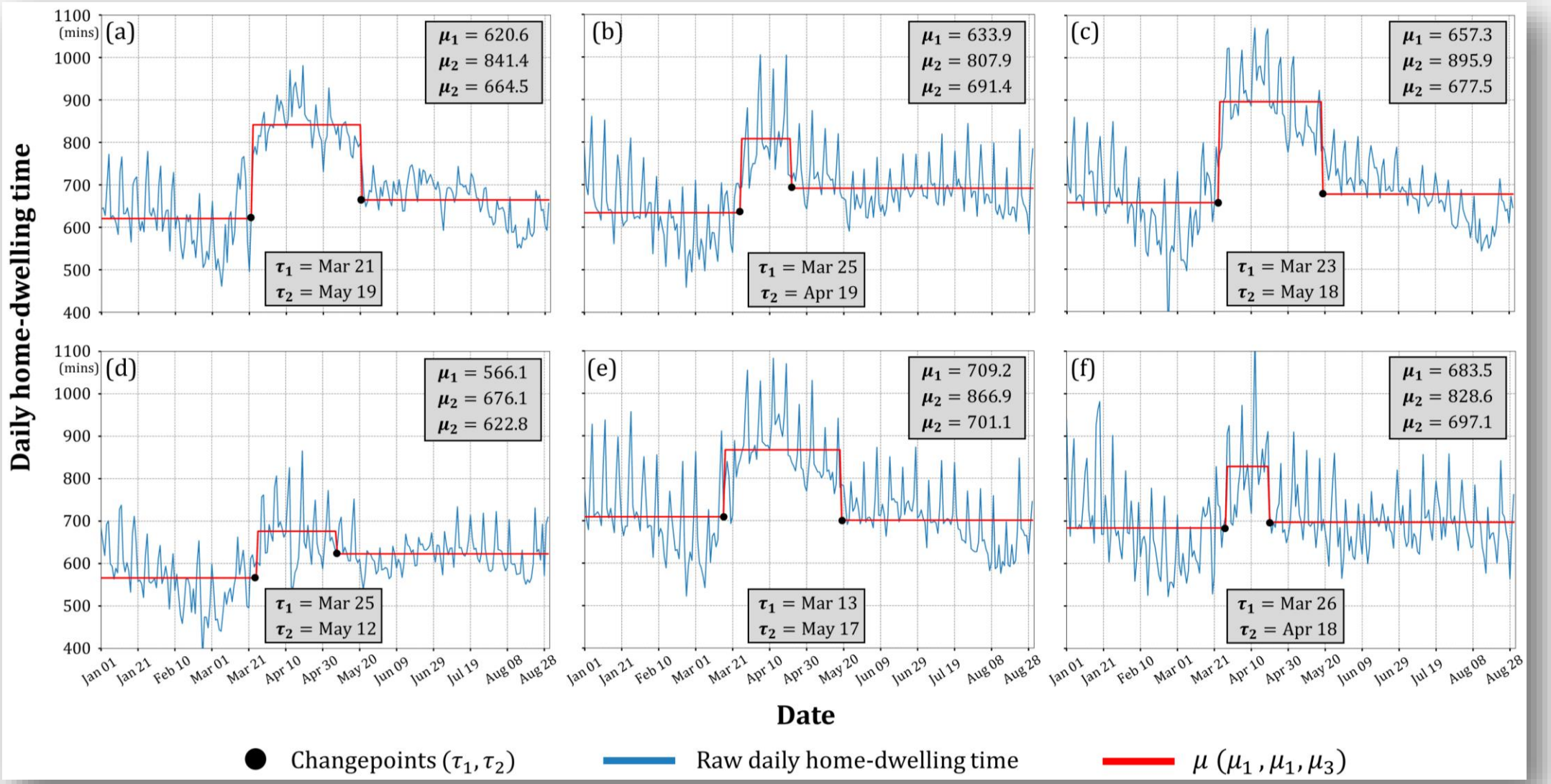
U.S. as a whole


Stage 1
Stage 2
Stage 3
Changepoints
Raw home-dwelling time
μ
$\mu \pm 3\sigma$
Daily home-dwelling time
1000 (mins)
900
800
700
600
500
$\mu_1 \sim \mathcal{N}(638.7, 7.6)$
$\mu_2 \sim \mathcal{N}(838.2, 9.7)$
$\mu_3 \sim \mathcal{N}(673.7, 6.6)$
$\tau_1 \approx$ Mar 22
$\tau_2 \approx$ May 17
Jan 01
Jan 21
Feb 10
Mar 01
Mar 21
Apr 10
Apr 30
May 20
Jun 09
Jun 29
Jul 19
Aug 08
Aug 28

U.S. States

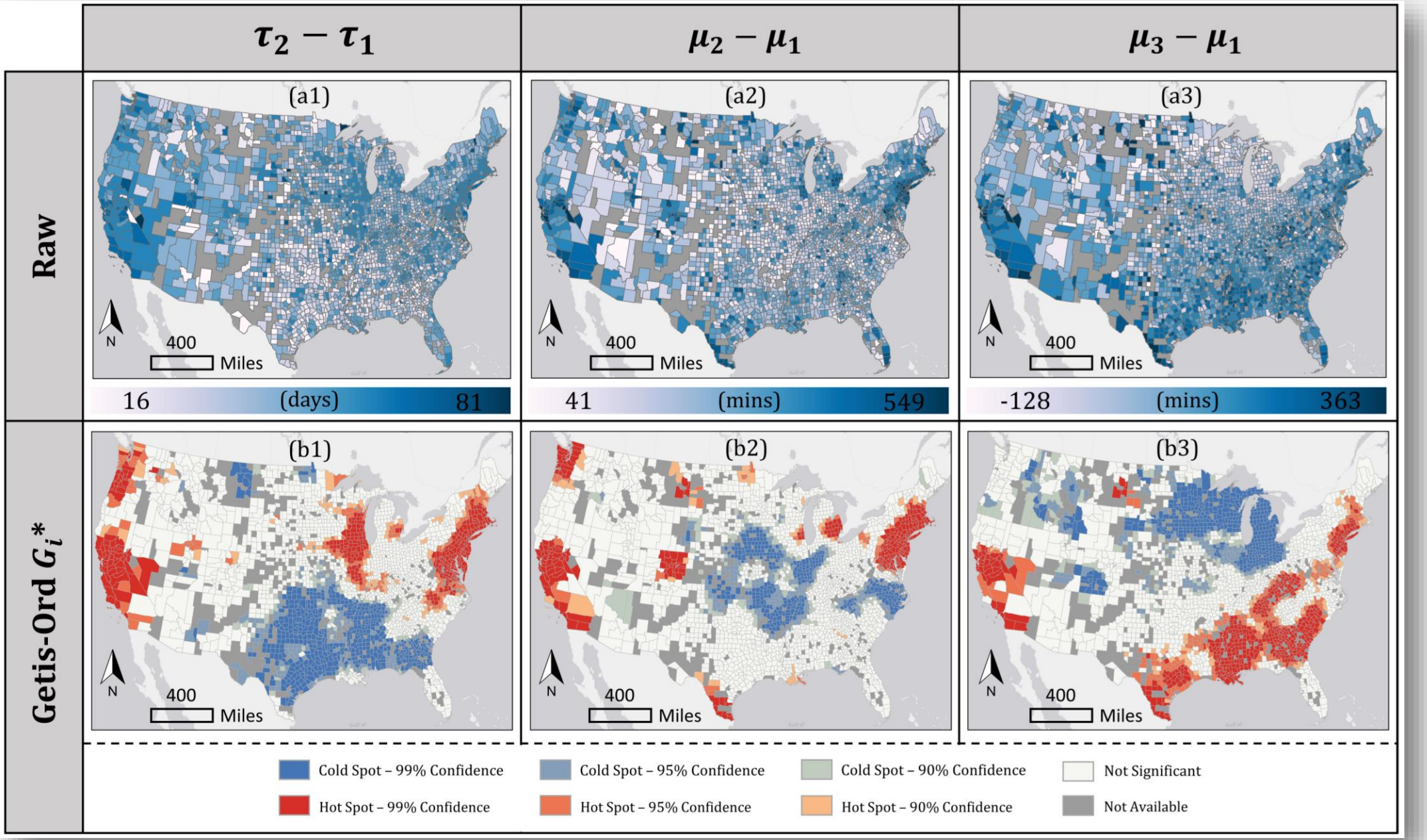

Daily home-dwelling time, derived stages, and stage related variables (μ_1, μ_2, μ_3, τ_1, τ_2) at the U.S. State level
AK $\mu_1 = 476.1$ $\mu_2 = 658.8$ $\mu_3 = 511.3$ τ_1 = Mar 23 τ_2 = May 13
ME $\mu_1 = 508.7$ $\mu_2 = 671.6$ $\mu_3 = 524.8$ τ_1 = Mar 25 τ_2 = May 17
WI $\mu_1 = 605.1$ $\mu_2 = 777.0$ $\mu_3 = 598.4$ τ_1 = Mar 22 τ_2 = May 18
VT $\mu_1 = 437.8$ $\mu_2 = 625.7$ $\mu_3 = 500.0$ τ_1 = Mar 25 τ_2 = May 12
NH $\mu_1 = 557.2$ $\mu_2 = 778.7$ $\mu_3 = 588.7$ τ_1 = Mar 24 τ_2 = May 18
WA $\mu_1 = 658.2$ $\mu_2 = 888.3$ $\mu_3 = 684.0$ τ_1 = Mar 22 τ_2 = May 19
ID $\mu_1 = 637.4$ $\mu_2 = 775.6$ $\mu_3 = 607.5$ τ_1 = Mar 24 τ_2 = May 17
MT $\mu_1 = 545.3$ $\mu_2 = 673.3$ $\mu_3 = 519.9$ τ_1 = Mar 25 τ_2 = May 13
ND $\mu_1 = 542.9$ $\mu_2 = 686.4$ $\mu_3 = 509.2$ τ_1 = Mar 25 τ_2 = May 13
MN $\mu_1 = 581.7$ $\mu_2 = 766.9$ $\mu_3 = 579.9$ τ_1 = Mar 24 τ_2 = May 18
IL $\mu_1 = 652.6$ $\mu_2 = 865.4$ $\mu_3 = 684.8$ τ_1 = Mar 20 τ_2 = May 18
MI $\mu_1 = 616.7$ $\mu_2 = 850.1$ $\mu_3 = 618.1$ τ_1 = Mar 22 τ_2 = May 18
NY $\mu_1 = 638.1$ $\mu_2 = 912.7$ $\mu_3 = 688.9$ τ_1 = Mar 21 τ_2 = May 19
MA $\mu_1 = 570.3$ $\mu_2 = 828.1$ $\mu_3 = 613.8$ τ_1 = Mar 22 τ_2 = May 19
OR $\mu_1 = 642.8$ $\mu_2 = 834.4$ $\mu_3 = 654.9$ τ_1 = Mar 23 τ_2 = May 19
NV $\mu_1 = 660.1$ $\mu_2 = 891.6$ $\mu_3 = 698.3$ τ_1 = Mar 20 τ_2 = May 18
WY $\mu_1 = 515.4$ $\mu_2 = 636.5$ $\mu_3 = 503.1$ τ_1 = Mar 30 τ_2 = May 16
SD $\mu_1 = 561.3$ $\mu_2 = 704.3$ $\mu_3 = 543.5$ τ_1 = Mar 25 τ_2 = May 17
IA $\mu_1 = 638.4$ $\mu_2 = 791.0$ $\mu_3 = 646.6$ τ_1 = Mar 24 τ_2 = May 18
IN $\mu_1 = 618.1$ $\mu_2 = 767.7$ $\mu_3 = 620.6$ τ_1 = Mar 24 τ_2 = May 16
OH $\mu_1 = 618.8$ $\mu_2 = 795.5$ $\mu_3 = 628.8$ τ_1 = Mar 22 τ_2 = May 16
PA $\mu_1 = 608.5$ $\mu_2 = 831.4$ $\mu_3 = 636.2$ τ_1 = Mar 21 τ_2 = May 18
NJ $\mu_1 = 681.8$ $\mu_2 = 1007.4$ $\mu_3 = 743.7$ τ_1 = Mar 21 τ_2 = May 19
CT $\mu_1 = 599.3$ $\mu_2 = 842.5$ $\mu_3 = 645.1$ τ_1 = Mar 21 τ_2 = May 19
RI $\mu_1 = 576.4$ $\mu_2 = 801.3$ $\mu_3 = 608.7$ τ_1 = Mar 23 τ_2 = May 19
CA $\mu_1 = 666.5$ $\mu_2 = 940.8$ $\mu_3 = 755.3$ τ_1 = Mar 20 τ_2 = May 19
UT $\mu_1 = 665.0$ $\mu_2 = 829.3$ $\mu_3 = 664.5$ τ_1 = Mar 20 τ_2 = May 17
CO $\mu_1 = 628.9$ $\mu_2 = 838.6$ $\mu_3 = 635.9$ τ_1 = Mar 24 τ_2 = May 16
NE $\mu_1 = 634.1$ $\mu_2 = 773.9$ $\mu_3 = 641.6$ τ_1 = Mar 24 τ_2 = May 17
MO $\mu_1 = 668.3$ $\mu_2 = 826.3$ $\mu_3 = 678.6$ τ_1 = Mar 20 τ_2 = May 17
KY $\mu_1 = 623.1$ $\mu_2 = 788.6$ $\mu_3 = 648.3$ τ_1 = Mar 24 τ_2 = May 17
WV $\mu_1 = 587.0$ $\mu_2 = 752.4$ $\mu_3 = 620.7$ τ_1 = Mar 24 τ_2 = May 11
VA $\mu_1 = 627.6$ $\mu_2 = 820.9$ $\mu_3 = 666.8$ τ_1 = Mar 23 τ_2 = May 19
MD $\mu_1 = 615.5$ $\mu_2 = 859.6$ $\mu_3 = 670.2$ τ_1 = Mar 21 τ_2 = May 19
DE $\mu_1 = 577.8$ $\mu_2 = 787.9$ $\mu_3 = 620.4$ τ_1 = Mar 23 τ_2 = May 19
AZ $\mu_1 = 655.5$ $\mu_2 = 833.1$ $\mu_3 = 699.4$ τ_1 = Mar 24 τ_2 = May 14
NM $\mu_1 = 560.0$ $\mu_2 = 714.3$ $\mu_3 = 591.2$ τ_1 = Mar 24 τ_2 = May 16
KS $\mu_1 = 651.1$ $\mu_2 = 811.9$ $\mu_3 = 668.6$ τ_1 = Mar 23 τ_2 = May 17
AR $\mu_1 = 653.1$ $\mu_2 = 770.7$ $\mu_3 = 678.8$ τ_1 = Mar 25 τ_2 = May 16
TN $\mu_1 = 634.1$ $\mu_2 = 785.9$ $\mu_3 = 666.2$ τ_1 = Mar 23 τ_2 = May 12
NC $\mu_1 = 625.1$ $\mu_2 = 785.4$ $\mu_3 = 649.9$ τ_1 = Mar 24 τ_2 = May 19
SC $\mu_1 = 635.9$ $\mu_2 = 785.4$ $\mu_3 = 662.1$ τ_1 = Mar 24 τ_2 = May 17
DC $\mu_1 = 463.1$ $\mu_2 = 599.2$ $\mu_3 = 497.3$ τ_1 = Mar 24 τ_2 = May 26
OK $\mu_1 = 670.0$ $\mu_2 = 829.6$ $\mu_3 = 693.0$ τ_1 = Mar 25 τ_2 = Apr 27
LA $\mu_1 = 619.0$ $\mu_2 = 793.0$ $\mu_3 = 665.8$ τ_1 = Mar 21 τ_2 = May 12
MS $\mu_1 = 630.2$ $\mu_2 = 801.5$ $\mu_3 = 692.0$ τ_1 = Mar 24 τ_2 = May 03
AL $\mu_1 = 679.9$ $\mu_2 = 830.1$ $\mu_3 = 715.6$ τ_1 = Mar 24 τ_2 = May 10
GA $\mu_1 = 662.8$ $\mu_2 = 850.4$ $\mu_3 = 706.0$ τ_1 = Mar 22 τ_2 = May 16
HI $\mu_1 = 591.6$ $\mu_2 = 863.1$ $\mu_3 = 721.2$ τ_1 = Mar 22 τ_2 = May 17
TX $\mu_1 = 665.1$ $\mu_2 = 864.4$ $\mu_3 = 719.1$ τ_1 = Mar 22 τ_2 = May 11
FL $\mu_1 = 638.7$ $\mu_2 = 828.9$ $\mu_3 = 673.4$ τ_1 = Mar 24 τ_2 = May 16
1400
1200
1000
800
600
400
200
0
(mins)
Daily home-dwelling time
Detected stages (μ)
Jan 01
Feb 19
Apr 09
May 29
Jul 18

U.S. Counties

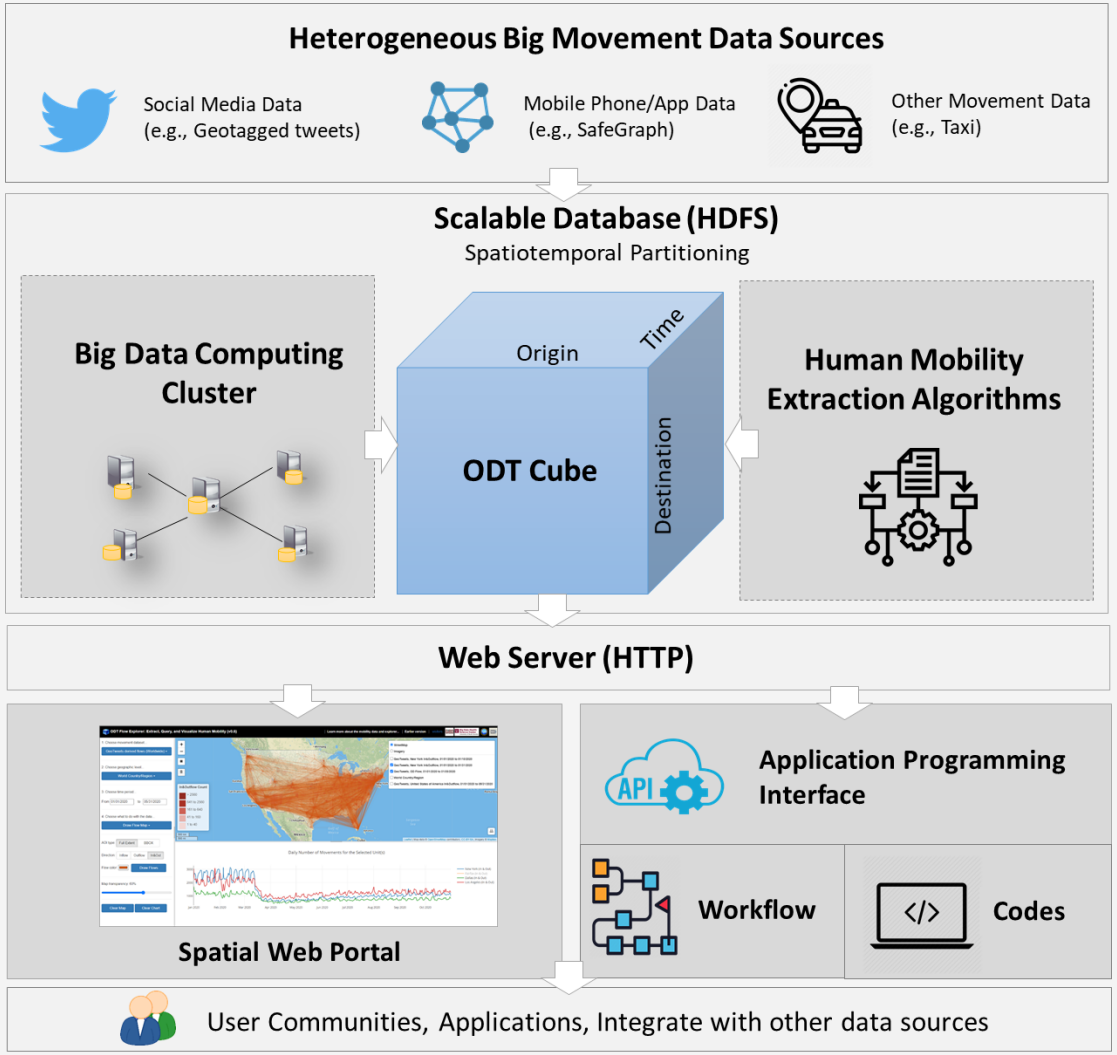


(a) Fulton County, Georgia; (b) Wayne County, Georgia (c) Pierce County, Washington; (d) Jefferson County, Arkansas; (e) Utah County, Utah; (f) Adaur County, Iowa.

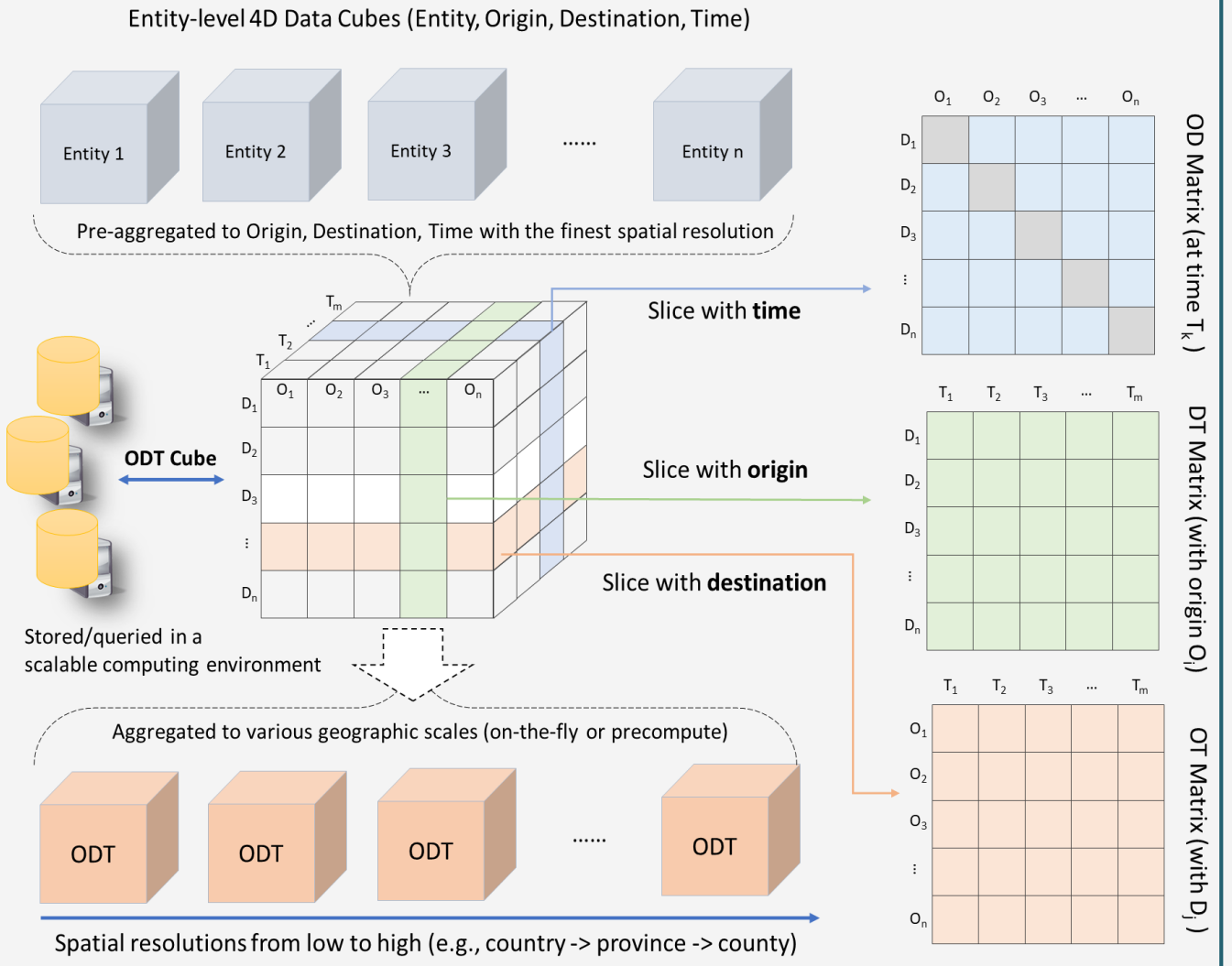
U.S. Counties


$\tau_2 - \tau_1$
$\mu_2 - \mu_1$
$\mu_3 - \mu_1$
Raw
Getis-Ord G_i^*
(a1)
(a2)
(a3)
(b1)
(b2)
(b3)
N
400
Miles
16
(days)
81
41
(mins)
549
-128
(mins)
363
Cold Spot – 99% Confidence
Cold Spot – 95% Confidence
Cold Spot – 90% Confidence
Not Significant
Hot Spot – 99% Confidence
Hot Spot – 95% Confidence
Hot Spot – 90% Confidence
Not Available

ODT Flow Platform

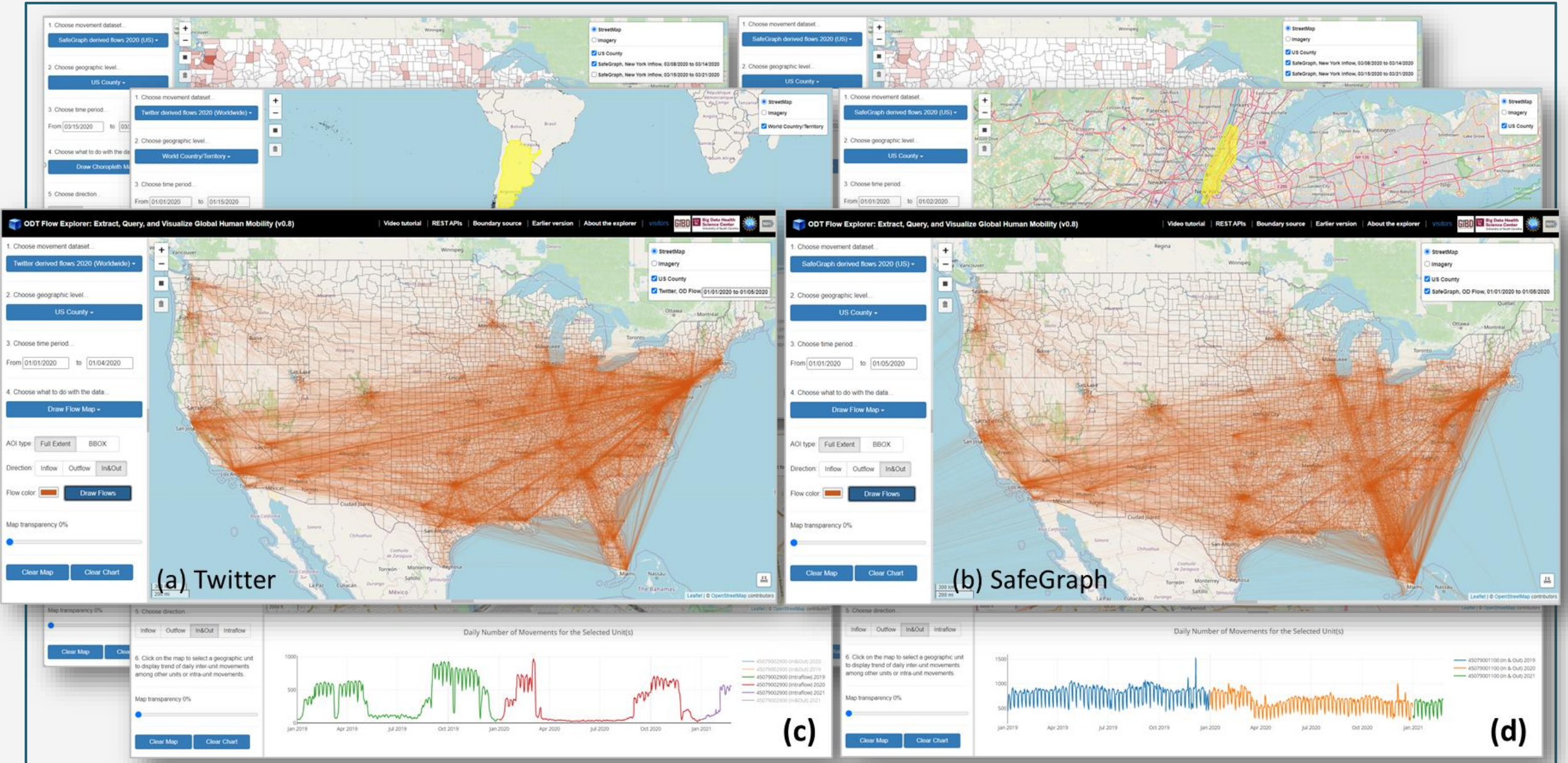


The system architecture of the ODT Flow platform



Origin-Destination-Time Cube

Li, Z., Huang, X., Hu, T., Ning, H., Ye, X., & Li, X. (2021). ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility. *arXiv preprint arXiv:2104.05040*.



Li, Z., Huang, X., Hu, T., Ning, H., Ye, X., & Li, X. (2021). ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility. *arXiv preprint arXiv:2104.05040*.

ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility Flows

http://gis.cas.sc.edu/GeoAnalytics/od.html

ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility Flows

http://gis.cas.sc.edu/GeoAnalytics/od.html

Zhenlong Li, Xiao Huang, Tao Hu, Huan Ning, Xinyue Ye, Xiaoming Li

Geoinformation and Big Data Research Lab, Department of Geography, University of South Carolina, Columbia, SC, USA

Department of Geosciences, University of Arkansas, Fayetteville, AR, USA

Center for Geographic Analysis, Harvard University, Cambridge, MA, USA

Department of Landscape Architecture & Urban Planning, Texas A&M University, TX, USA

Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, USA

Li, Z., Huang, X., Hu, T., Ning, H., Ye, X., & Li, X. (2021). ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility. *arXiv preprint arXiv:2104.05040*.

Huang, X., Li, Z., Jiang, Y., Ye, X., Deng, C., Zhang, J., & Li, X. (2021). The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. International Journal of Digital Earth, 1-19.

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.

Huang, X., Li, Z., Lu, J., Wang, S., Wei, H., & Chen, B. (2020). Time-series clustering for home dwell time during COVID-19: what can we learn from it?. ISPRS International Journal of Geo-Information, 9(11), 675.

Huang, X., Lu, J., Gao, S., Wang, S., Liu, Z., & Wei, H. (2021). **Staying at home is a privilege**: evidence from fine-grained mobile phone location data in the U.S. during the COVID-19 pandemic. Annals of the American Association of Geographers. doi: 10.1080/24694452.2021.1904819. (PDF) Staying at home is a privilege: evidence from fine-grained mobile phone location data in the U.S. during the COVID-19 pandemic.

Huang, X., Xu, Y., Liu, R., Wang, S., Wang, S., Zhang, M., Kang, Y., Zhang, Z., Gao, S., Li, Z. Exploring home-dwelling time patterns during the COVID-19 pandemic via Bayesian inference: policy compliance, spatial disparity, and exposed inequity. (To be submitted)

Li, Z., **Huang, X.**, Hu, T., Ning, H., Ye, X., & Li, X. (2021). ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility. arXiv preprint arXiv:2104.05040.

Li, Z., **Huang, X.**, Ye, X., Jiang, Y., Yago, M., Ning, H., ... & Li, X. (2021). Measuring Place Connectivity Using Big Social Media Data. arXiv preprint arXiv:2102.03991.

Hu, T., Wang, S., She, B., Zhang, M., **Huang, X.,** Cui, Y., ... & Li, Z. (2021). Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges (Under Review).

Computational Urban Science

Editors-in-Chief: Hui Lin, Xinyue Ye

Special Issue:

Promoting Urban Computational Paradigms with Shareable Data, Models, Tools, and Frameworks

Guest Editors:

- Xiao Huang, University of Arkansas (xh010@uark.edu)
- Alexander Hohl, University of Utah (alexander.hohl@geog.utah.edu)
- Zhenlong Li, University of South Carolina (zhenlong@mailbox.sc.edu)

Springer

Deadline: Dec 31, 2021

Special Issue "Spatial Analytics for COVID-19 Studies"

- Special Issue Editors
- Special Issue Information
- Keywords
- Published Papers

International Journal of *Environmental Research and Public Health*

A special issue of *International Journal of Environmental Research and Public Health* (ISSN 1660-4601). This special issue belongs to the section "Environmental Health".

Deadline for manuscript submissions: **1 May 2022**.

Share This Special Issue

Special Issue Editors

Dr. Tao Hu E-Mail Website
Guest Editor
Center for Geographic Analysis, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138, USA
Interests: spatial modelling; geospatial big data analysis; health geography; urban crime analysis

Dr. Zhenlong Li E-Mail Website SciProfiles
Guest Editor
Geoinformation and Big Data Research Laboratory (GIBD), Department of Geography, University of South Carolina, Columbia, SC 29208, USA
Interests: GIScience; spatial computing; geospatial big data; social media analytics; CyberGIS
Special Issues and Collections in MDPI journals

Dr. Xiao Huang E-Mail Website SciProfiles
Guest Editor
Department of Geosciences, University of Arkansas, Fayetteville, AR 72701, USA
Interests: geospatial analysis; remote sensing; big data analytics; geovisualization; data fusion

Deadline: May 1, 2022

UNIVERSITY OF ARKANSAS

THANK YOU / QUESTIONS ?

Xiao Huang (xh010@uark.edu)
Department of Geosciences
University of Arkansas
Personal website: www.xiaohuang116.com

May 07, 2021