

Spatial Big Data Analysis to Understand COVID-19 Effects on MN

By Yan Li, Arun Sharma, Shashi Shekhar

Presenter: Shashi Shekhar

McKnight Distinguished University Professor, Univ. of Minnesota

shekhar@umn.edu

http://www.spatial.cs.umn.edu/Project/covid_19.html

Forum: Impacts of the COVID-19 Pandemic on Minnesota's Traffic and Transit Networks

Center for Transportation Studies, University of Minnesota, June 23rd 2020



UNIVERSITY OF MINNESOTA
Driven to Discover™

SPATIAL
COMPUTING
SHASHI SHEKHAR AND
PAMELA VOLO



THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES

Spatial Databases
A TOUR



Shashi Shekhar • Sanjay Chandra

SPRINGER
REFERENCE

Shashi Shekhar
Hai Xiong
Xun Zhou
Editors

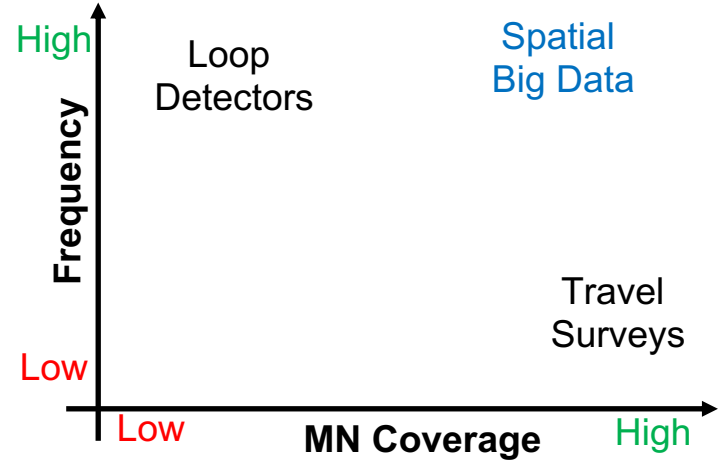
Encyclopedia
of GIS

Second Edition

Springer

Why Spatial Big Data?

- Challenges
 - COVID-19 impacts entire state
 - Frequent policy interventions
- Limitations of State of the Art
 - Loop detectors: 3 MN cities (major roads)
 - Travel surveys are infrequent
- Proposed Approach: **Spatial Big Data**
 - Mobile Phone location traces
 - Ex. Streetlight, Cuebiq, SafeGraph
 - Caution: privacy, bias, volume, velocity, variety
- **Note:** Complements does not replace other datasets!



Spatial Big Data

- Location traces
 - 2 billion GPS receivers today (7 billion by 2022)
 - Reference clock for telecom, banks, ...
 - Help understand Spatio-temporal patterns of life
- Other:
 - (Nano-)Satellite Imagery, ...



The World Economy
Runs on GPS. It Needs a
Backup Plan

Bloomberg Businessweek

July 25, 2018, 4:00 AM CDT

ENSURING RESOURCE AVAILABILITY

Advanced technology, including many types of Earth information, will unlock up to **\$1.6 trillion** in economic savings for energy generation and use by 2035.

Satellite observations can also help ensure water availability, which is particularly important to the 20% of the world now living in areas of water scarcity.

McKinsey Global Institute

The study estimates that the use of **personal location data** could save consumers worldwide more than **\$600 billion** annually by **2020**. Computers determine users' whereabouts by tracking their mobile devices, like cellphones.

The New York Times

Published: May 13, 2011

SafeGraph Overview

- MN Coverage:
 - 294,014 devices
 - 73,548 Points of Interests (across 261 categories)
 - 4,107 Census Block Groups (out of 4,111)
- Frequency:
 - Raw: periodic (location) pings from anonymous mobile devices
 - Aggregate: Hourly, Daily, Weekly, Monthly, ...
 - Recency: a few days
- Multiple datasets:
 - Social distancing: average range and time-at-home by census block group
 - Weekly pattern: POI visits by hour, day, and week
 - Monthly pattern
 - ...

Social Distancing dataset

- Number of devices:
 - 294,014 (MN)
 - 154,467 (Twincities)
- Schema Summary
 - Details: <https://docs.safegraph.com/docs/social-distancing-metrics>

Name	Description	Type
completely_home_device_count	the number of devices which did not leave their home	Int
median_home_dwell_time	Median dwell time at home	Int
bucketed_home_dwell_time	Key: range of time; Value: device count	{Str: Int}
at_home_by_each_hour	List of device count in each hour in a day	[Int]
part_time_work_behavior_devices	the number of devices that spent one period of between 3 and 6 hours at one location other than their home	Int
full_time_work_behavior_devices	the number of devices that spent greater than 6 hours at a location other than their home	Int

Name	Description	Type
<u>origin_census_block_group</u>	The unique 12-digit FIPS code for the census block group.	Str
date_range_start	Start time for measurement period	Str
date_range_end	End time for measurement period	Str
device_count	Number of devices during the date range whose home is in this census_block_group.	Int
distance_traveled_from_home	Median distance (in meters) traveled from home during the time period	Int
bucketed_distance_traveled	Key: range of meters; Value: device count.	{Str: Int}
median_dwell_at_bucketed_distance_traveled	Key: range of meters; Value: time	{Str: Int}

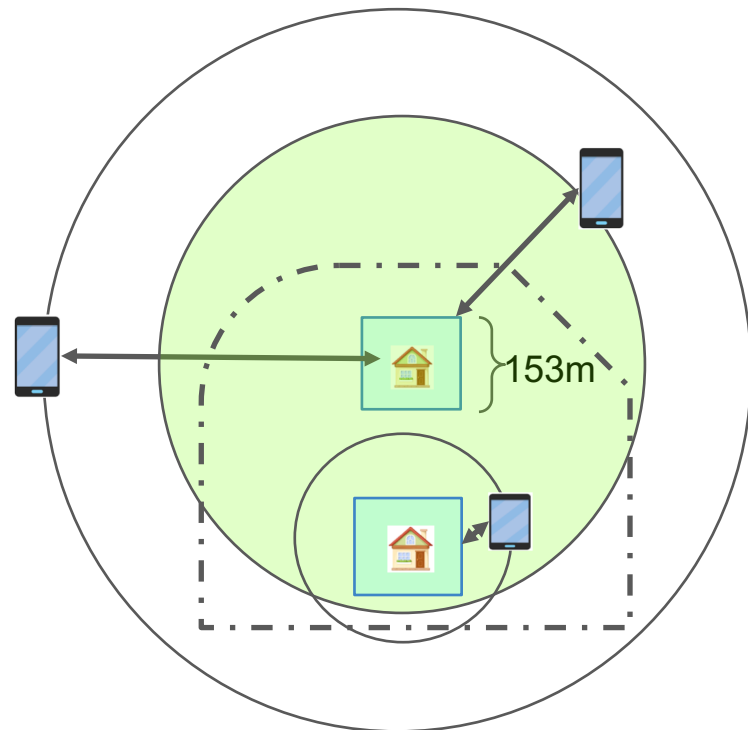
Median Range By Census Block Group

`distance_traveled_from_home`

Median distance (in meters) traveled from home during the time period

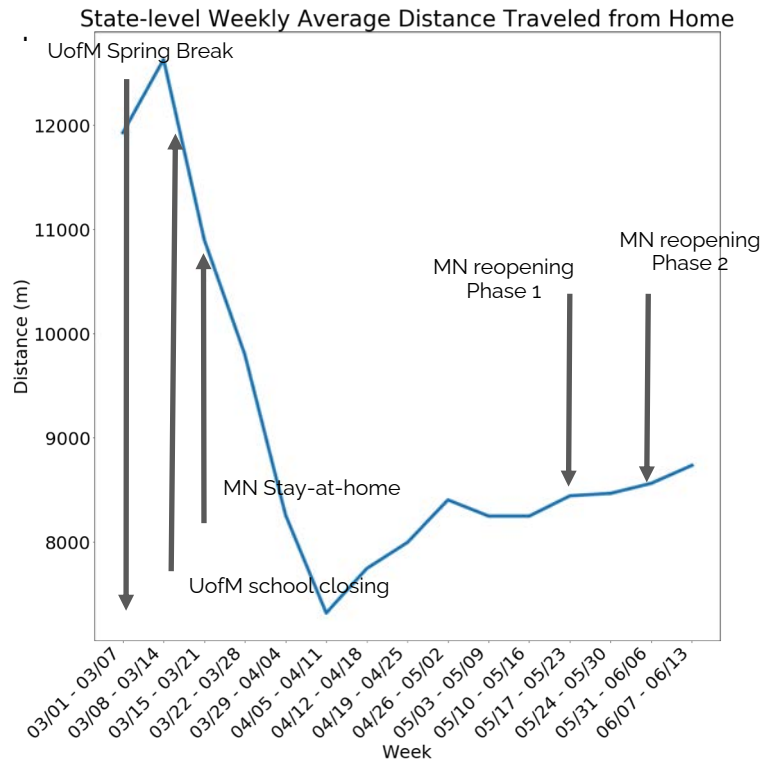
Int

- "Home" geohash (153m x 153m):
 - Common nighttime location over 6-weeks
 - Relocation effects, e.g., Spring break
- Aggregation by census block groups (CBGs)
 - Data suppressed for CBG with few devices.
- Differential privacy similar to Census 2020
 - applied to all metrics except `device_count`
 - Protects individual information

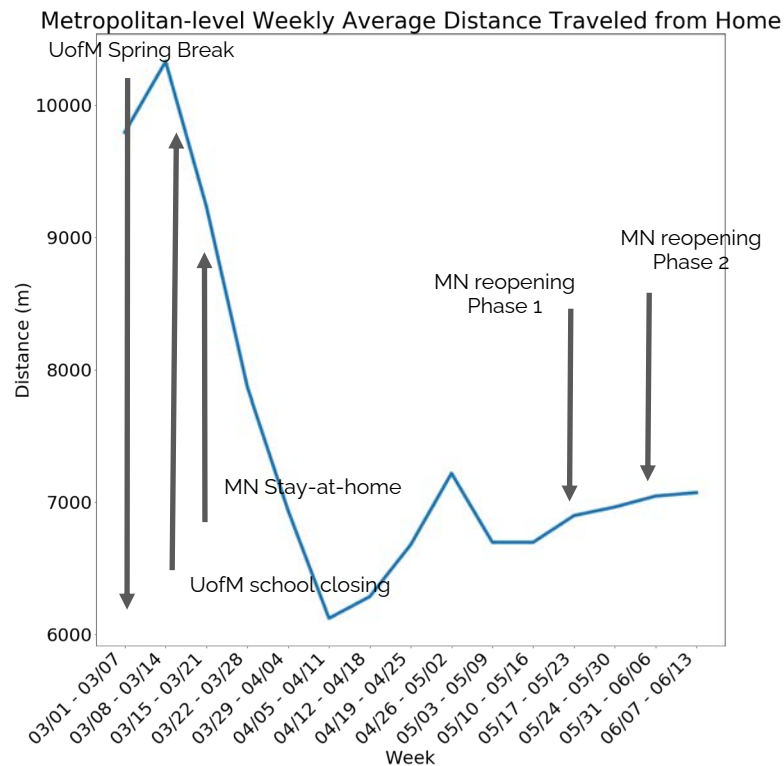


Analysis of MN Daily Range

- **Trends:** Sharp Decline in March, slow recovery starting mid-April
- **Note:** Data reporting changed on 5/10)



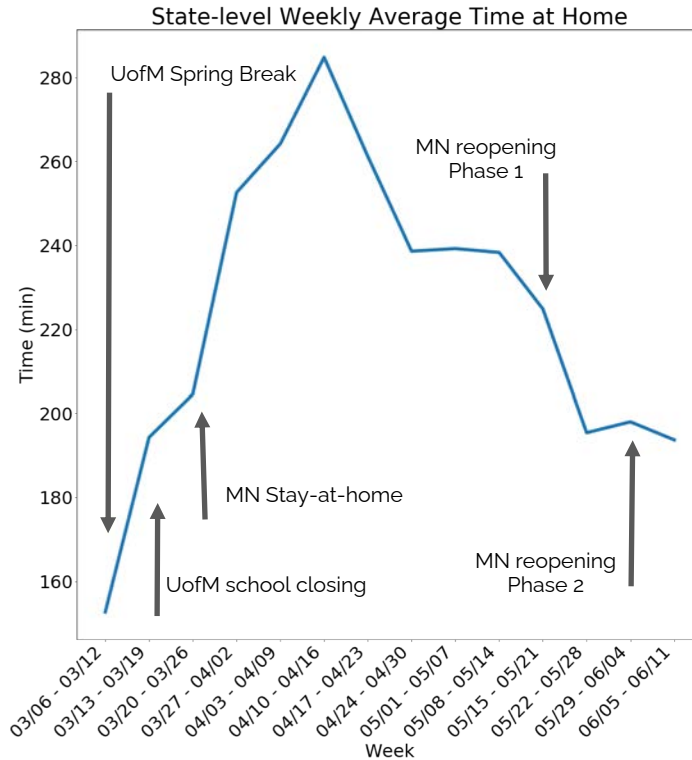
State of Minnesota



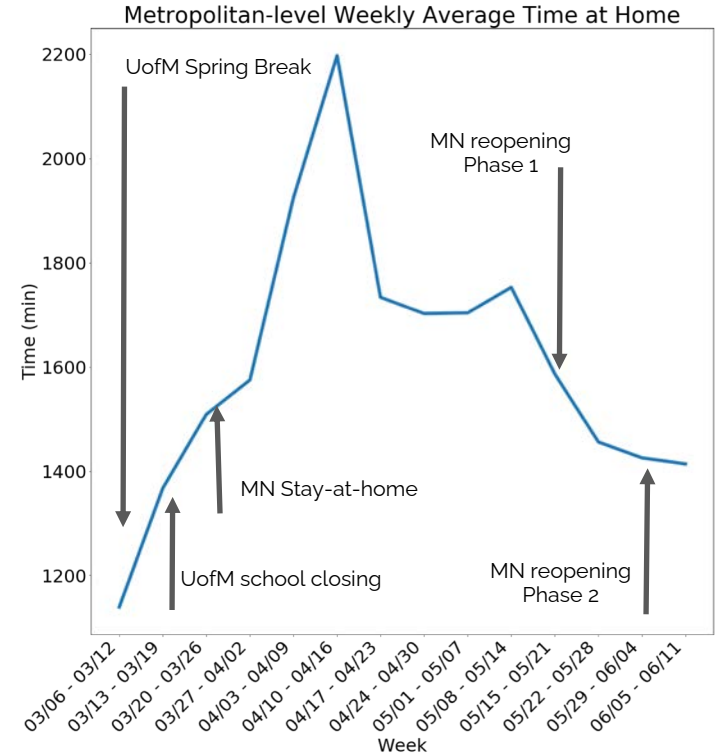
Twincities

Daily average time at home

- **Trends:** Increases in March then declines after mid-April
- Q? Where were people going after mid-April? POI visit data



State of Minnesota



Twincities

SafeGraph Weekly Pattern Dataset

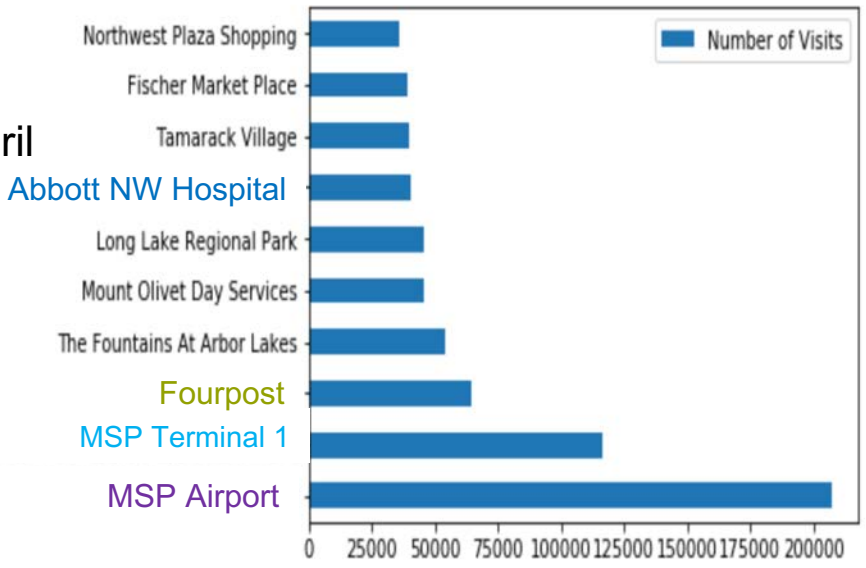
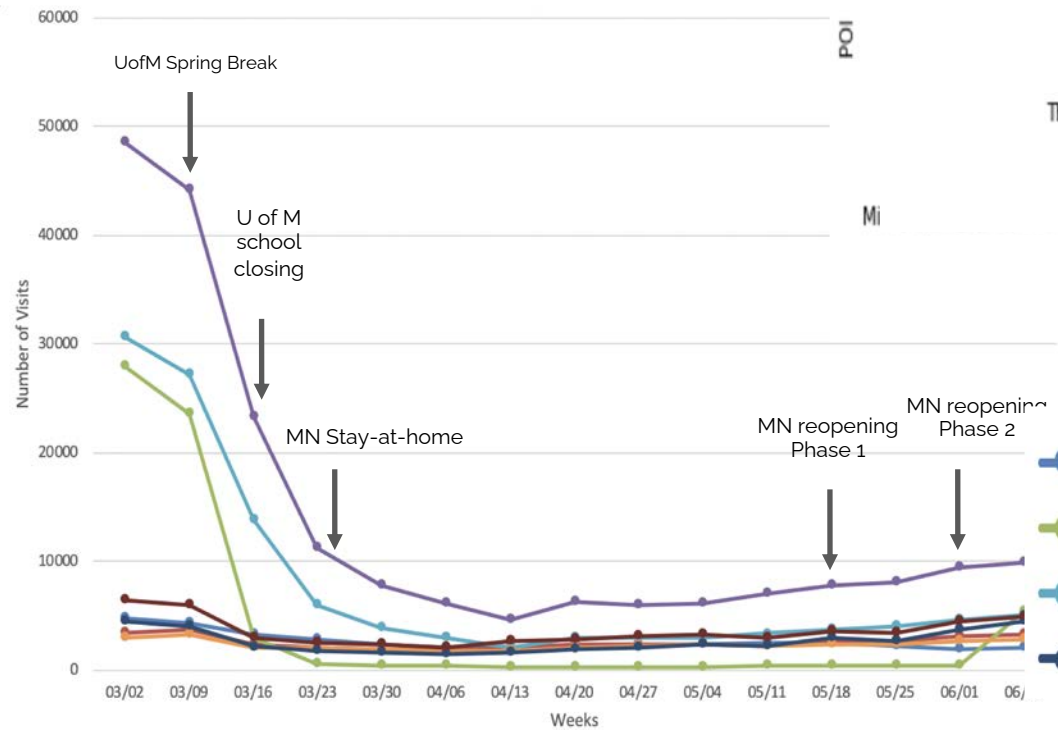
- Summary: MN: 73,548 POIs across 261 NAICS Code
- Schema Summary (Details: <https://docs.safegraph.com/docs/weekly-patterns>)

Name	Description	Type
<u>safegraph_place_id</u>	Unique ID tied to POI	Str
location_name	Name of POI	Str
street_address	Address of POI	Str
city	City where POI is located	Str
region	State or Territory	Str
postal_code	Zip Code	Str
brands	Name of the Business	List
naics_code	Code Describing Business	Int
date_range_start	Start Time in ISO 8601	Str
date_range_end	End Time in ISO 8601	Str
raw_visit_counts	Number of Visits	Int
raw_visitor_counts	Number of uniques visitors	Int

Name	Description	Type
visits_by_day	Visits by each day	Int
visits_by_each_hour	Visits by each hour	Int
visitor_home_cbgs	Number of visitors whose home is in that census block group	{Str: Int}
visitor_country_of_origin	Country to Visitors	{Str: Int}
distance_from_home	Median distance from home cbgs travelled by visitors.	Int
median_dwelling	Median Dwell Time	Double
bucketed_dwelling_time	Key: Minutes, Value: Visits	{Str: Int}
related_same_day_brand	Key: Brands, Value: Visits (by day)	{Str: Int}
related_same_week_brand	Key: Brands, Value: Visits (by week)	{Str: Int}
device_type	Key: Android or iOS, Value: Count	{Str: Int}
iso_country_code	2 letter country code	Str

Most Frequented POIs in MN

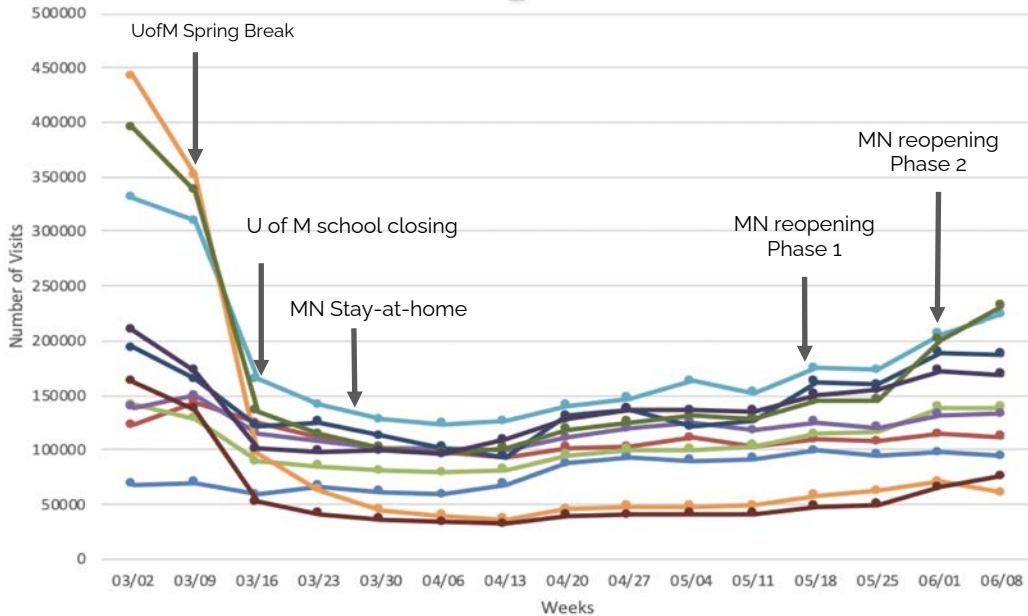
- Minnesota, March 2nd – June 8th, 2020
- **Trends:** March crash, some recovery mid-April



- Abbott NW Hospital
- Fischer Market Place
- Fourpost
- MSP Airport
- MSP Terminal 1
- Northwest Plaza Shopping
- Tamarack Village
- The Fountains At Arbor Lakes

Most Frequented Business Categories

- Minnesota, March 2nd – June 8th, 2020
- **Trends:** Large drops in March (except **Groceries**, **Hardware**,)
 - Increase in April (even **Schools** to return accessories) **Fitness and Sports** (late May)

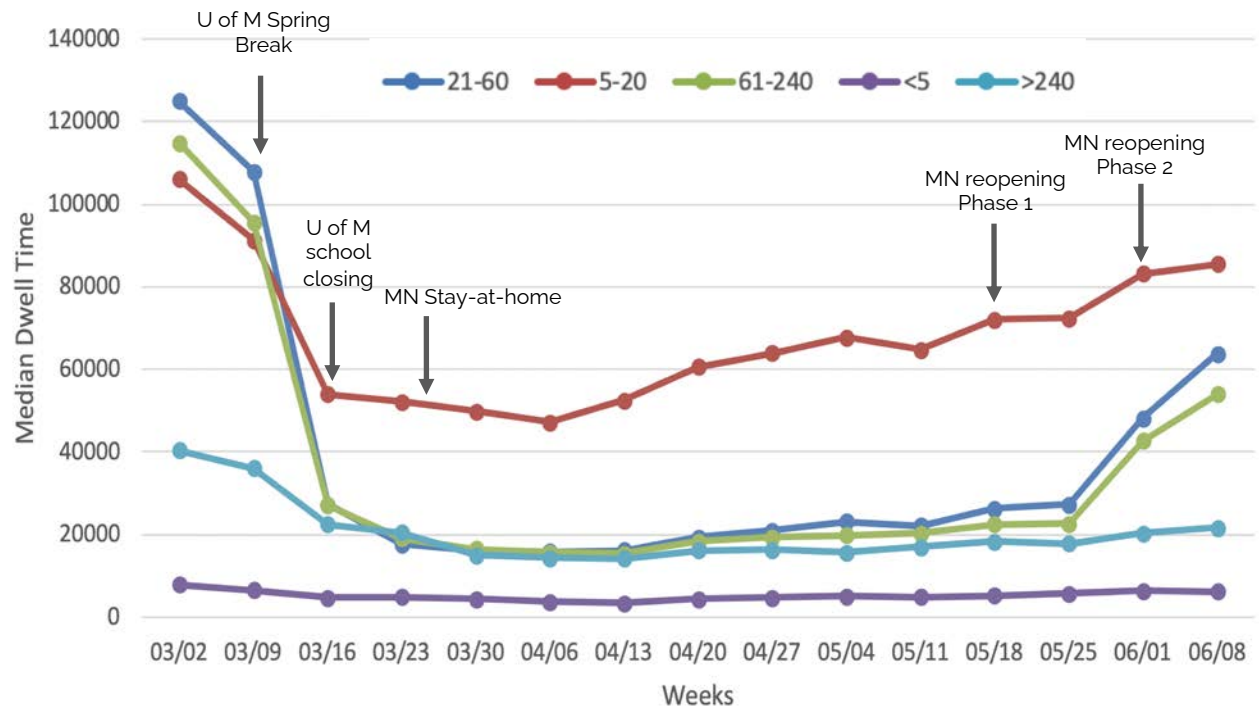


NAICS Code	Meaning
444130	Hardware Stores
445110	Supermarkets, Groceries
447110	Gasoline Station with Convenience Store
452319	Other General Merchandise Store
531120	Lessors of Non-Residential Buildings (Malls)
611110	Elementary and Secondary Schools
712190	Nature Parks and Other Similar Institutions
713940	Fitness and Recreational Sports Centers
722511	Full-Service Restaurants
722513	Limited Service Restaurants



Analyzing Restaurants Visits

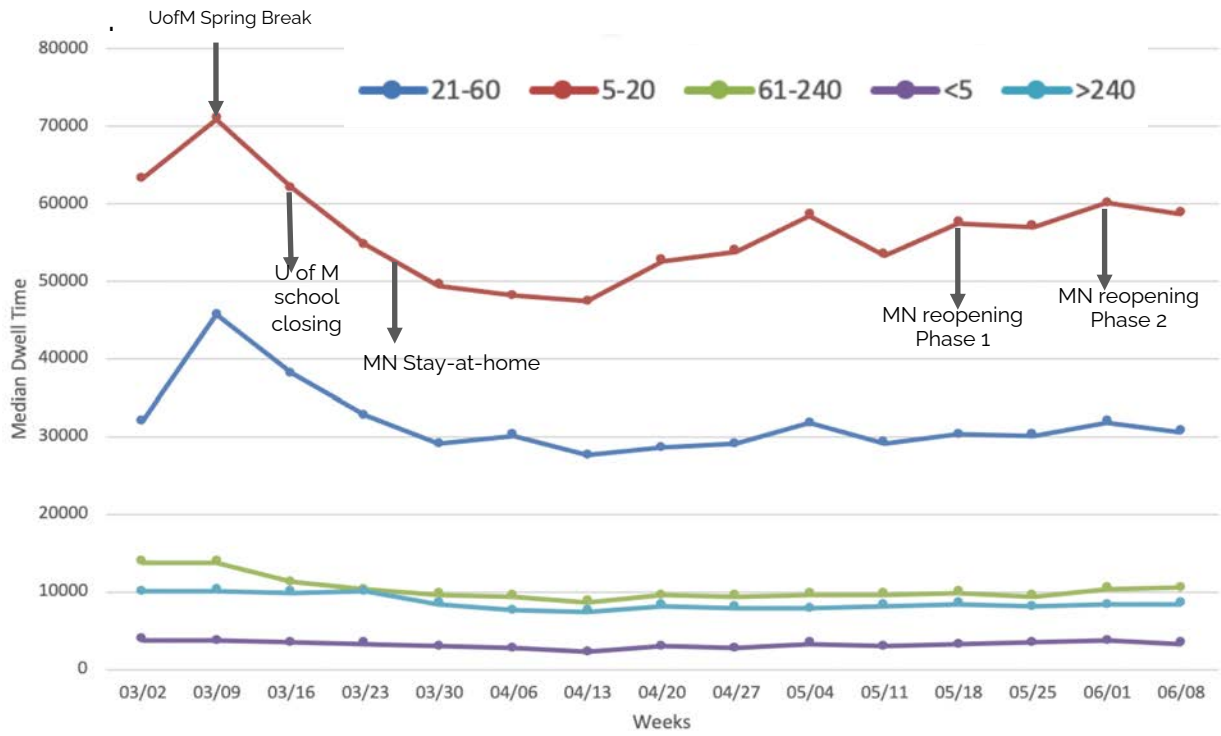
- Minnesota, March 2nd – June 8th, 2020
- **Trends:** After Mid March, **short (5-20 min)** visits dominate. ? Pickup and deliveries
 - **Longer** visits drop a lot, but short visits drop less and start to recover.



Dwell Time	Total Visits
<5	65741
5-20	854879
21-60	465689
61-240	427158
>240	264283

Analyzing Grocery Store Visits

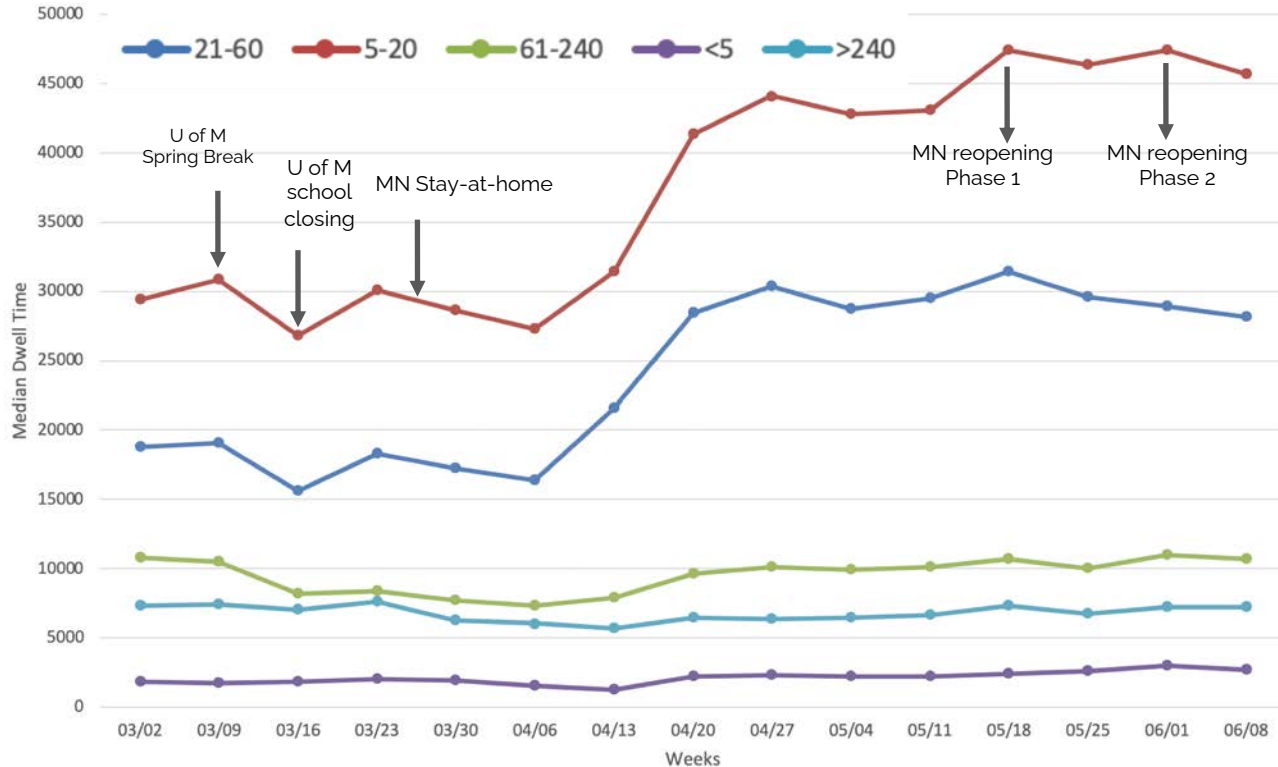
- Minnesota, March 2nd – June 8th, 2020
- Trends: Early March bump, **More Short (< 20 mins) trips** – ?pickup/delivery



Dwell Time	Total Visits
<5	40606
5-20	729132
21-60	414656
61-240	134546
>240	112405

Analyzing Hardware-Store Visits

- Minnesota, March 2nd – June 8th, 2020
- Trends: No March dip, **Short** and **medium** (5-60 min) visits increase starting April.



Dwell Time	Total Visits
<5	26253
5-20	469256
21-60	304960
61-240	121253
>240	87480

Summary & Next Steps

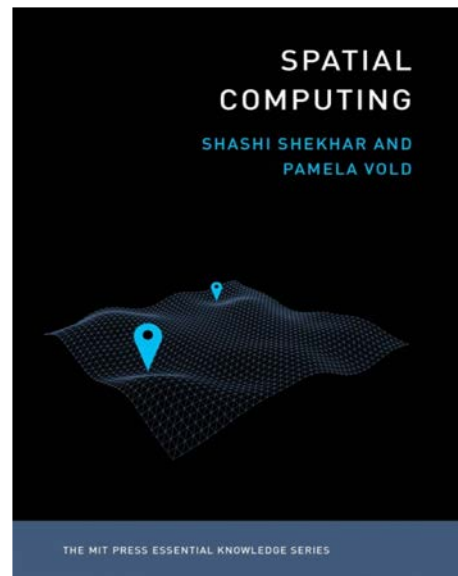
- Summary

- Spatial Big Data (SBD) quantifies COVID-19 impacts
- SBD supplements, not replaces, other mobility data
- New insights, e.g., restaurants, hardware stores, ...

- Next Steps

- Probe data quality, e.g., selection bias
- Link to census data to compare across census tracts
- Explore trends in part-time and full-time work
- Conceptual data model for flexible querying
- Spatial Data Mining Algorithms to automate routine tasks

- More: http://www.spatial.cs.umn.edu/Project/covid_19.html



Acknowledgements

- National Science Foundation (Smart & Connected Communities Grant)
- Data Provider: SafeGraph
- MIT media lab: Esteban Moro Egido and Alex Pentland
- Metropolitan Council: Ashley Asmus
- State: MnDOT, MMB, MN DHS
- University of Minnesota
 - Center for Transportation Studies: Laurie McGinnis
 - School of Public Health: Eva Enns, Shalini Kulasingam, Kelly Searle
 - U-Spatial: Len Kne, Adam Null
 - Dept. of Industrial and Systems Eng.: Ankur Mani