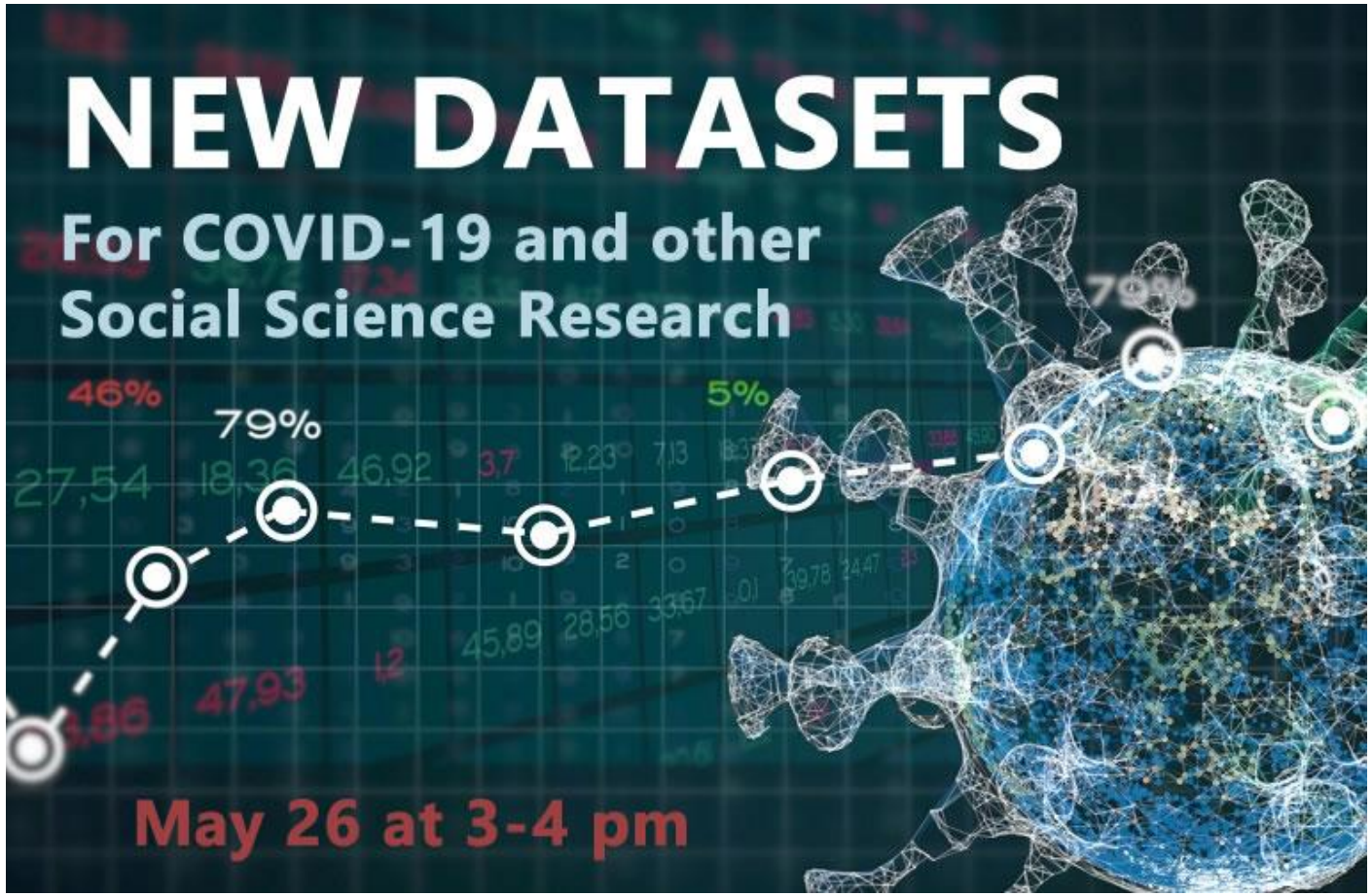



**Panelists:**

**Barbara Esty**, Data Librarian, Marx Library

**Eli Fenichel**, Knobloch Family Professor of Natural Resource Economics, F&ES

**Anton Gollwitzer**, PhD Student, Psychology

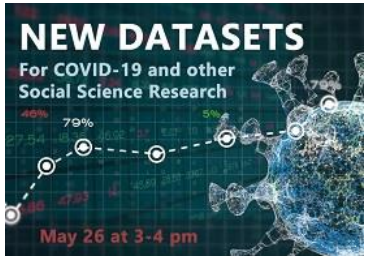
**Sara Gottlieb-Cohen**, Statistical Support Services Manager, Marx Library

**Roy Lederman**, Assistant Professor of Statistics and Data Science

**Cormac O'Dea**, Assistant Professor of Economics

**Moderator: Alan Gerber**, Dean of the Social Science Division of the Faculty of Arts and Science, the Charles C. and Dorathea S. Dilley Professor of Political Science, and incoming Director of the Institution for Social and Policy Studies

# New Datasets for COVID-19 and other Social Science Research

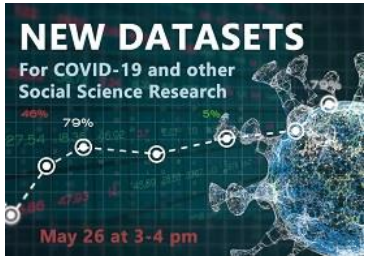



May 26, 2020

## Meeting logistics

- Send questions to limor.peer@yale.edu. The chat function, audio, and video will be disabled for this meeting.
- The event will be recorded. We will provide a link so that you can view the session later.
- All information from the session will be posted here: https://isps.yale.edu/new-datasets

# New Datasets for COVID-19 and other Social Science Research




May 26, 2020

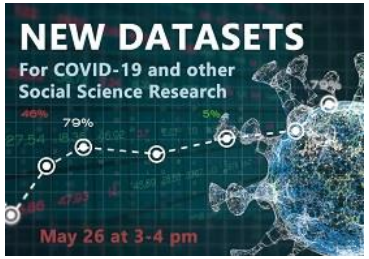
## Today's presenters:

- **Barbara Esty**, Data Librarian, Marx Library
- **Eli Fenichel**, Knobloch Family Professor of Natural Resource Economics, F&ES
- **Anton Gollwitzer**, PhD Student, Psychology
- **Sara Gottlieb-Cohen**, Statistical Support Services Manager, Marx Library
- **Roy Lederman**, Assistant Professor of Statistics and Data Science
- **Cormac O'Dea**, Assistant Professor of Economics

More information: https://isps.yale.edu/new-datasets

Yale

# New Datasets for COVID-19 and other Social Science Research



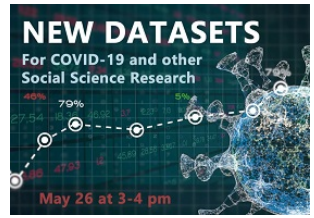

May 26, 2020

## Agenda:

- **Homebase** employment patterns in US small businesses (Cormac O'Dea)
- **Unacast Social Distancing Scoreboard** data (Anton Gollwitzer)
- **SafeGraph** cell-phone mobility data (Eli Fenichel, Roy Lederman)
- **Yale research data support** (Barbara Esty, Sara Gottlieb-Cohen)

More information: https://isps.yale.edu/new-datasets

Yale

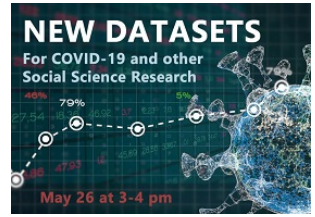



# Homebase Data

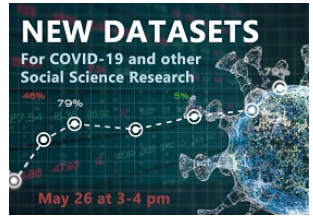
Cormac O'Dea

26 May 2020

# Homebase




- **Homebase** is company that provides scheduling and timeclock software to firms with hourly paid employees
  - Basic version free to firms, enhanced version billed
  - Services over 50,000 firms with over 450,000 employees
- They are making their data easily available to researchers
  - `https://joinhomebase.com/data/covid-19/`
  - If you're interested, email me `cormac.odea@yale.edu`

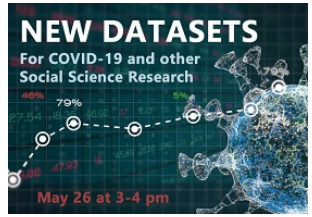



# What is in the data?

Think of this as a dataset of shifts

| Area | Variable | Description |
|---|---|---|
| Date | Event Date | Time clock event |
| | Job Created Date | Date the job was created |
| | Job Archived Date | Date the job ended |
| | Location Created Date | Date the location was created |
| Hours and Wages | Hourly wage rate | |
| | Hours worked | |
| Identifiers | Company Id | Firm identifier |
| | Location Id | Some firms have different locations |
| | User Id | Worker/Individual identifier |
| Location of firm | State | State code |
| | MSA | Metropolitan Statistical Area |
| | Zip | Zip code |
| | County | County Code |
| Firm/Employee Characteristics | Industry | Broad categories of industry |
| | Manager Indicator | |

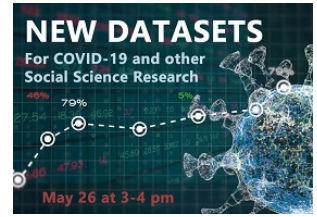
# What can we do with this data




The data . . .

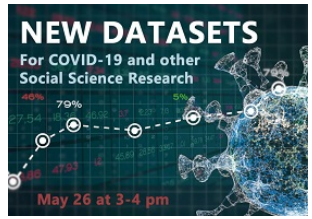
1. . . . allows a tracking of labor market conditions at a fine geographic level
2. . . . can be linked (by geography) with data on other conditions (health, mobility, crime etc)
3. . . . is timely: the data for yesterday (Monday 25) will have been uploaded by now.

# Note 1: Coverage



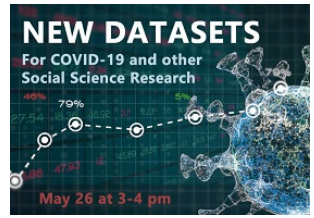

- This is not a dataset which is nationally-representative of firms or employees
  - Firms: predominantly small and medium sized businesses
  - Employees: almost all hourly-paid and predominantly low-paid

# Note 2: Following Individuals



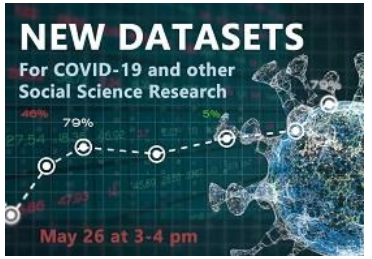

- We can follow individuals only to the extent that they remain working for a firm that uses Homebase
- What does a persistent fall in observed hours worked mean for an individual or group of individuals?
  - In March-May 2020 – that they're not working anywhere seems like a reasonable assumption for most
  - As firms generally start to open, this seems like a stronger assumption

# Our Plans




1. Governor's Office have asked us whether we can use this to provide timely evidence on what is happening in CT labor market
2. Plan to make available trends in aggregates and distributional impact through Tobin Center website
3. Medium term plan to look at association between these (local) labor market conditions other local characteristics (crime, health etc.)

If interested in accessing or using the data - please get in touch (`cormac.odea@yale.edu`)



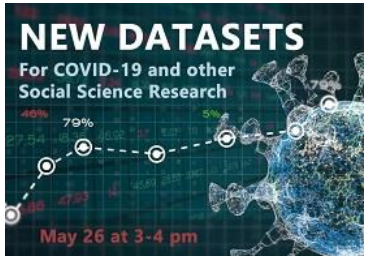

May 26, 2020

# Using Geo-Location Tracking Data in Research on COVID-19

*Anton Gollwitzer*

Contact: anton.gollwitzer@yale.edu

# Geo-Location Data



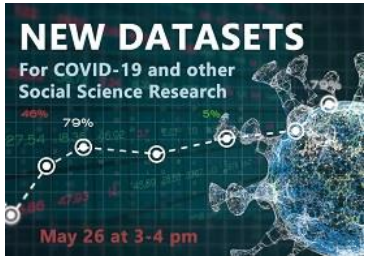

May 26, 2020

-Location data of millions of people as tracked by their smartphones over time

-275 million smartphone users in United States (2020)

-Can be aggregated at the state or county level for anonymization

# Geo-Location Data Sources




May 26, 2020

-Unacast

-Safegraph

-Descartes Lab

-Google Mobility Data*

-Apple Mobility Data*

*Can access without request but harder to work with (missing data and county names instead of county FIPS codes)
*Include foreign countries

unacast.

Search for state and county

Last updated human mobility data: May 25, 2020, 11:46:29 AM
Last updated COVID-19 cases: May 25, 2020, 3:55:01 AM
For more details on data accuracy please refer to our methodology section.
Explore this report in Tableau
unacast-covid-scoreboard@1.1.0-9b3fdc0

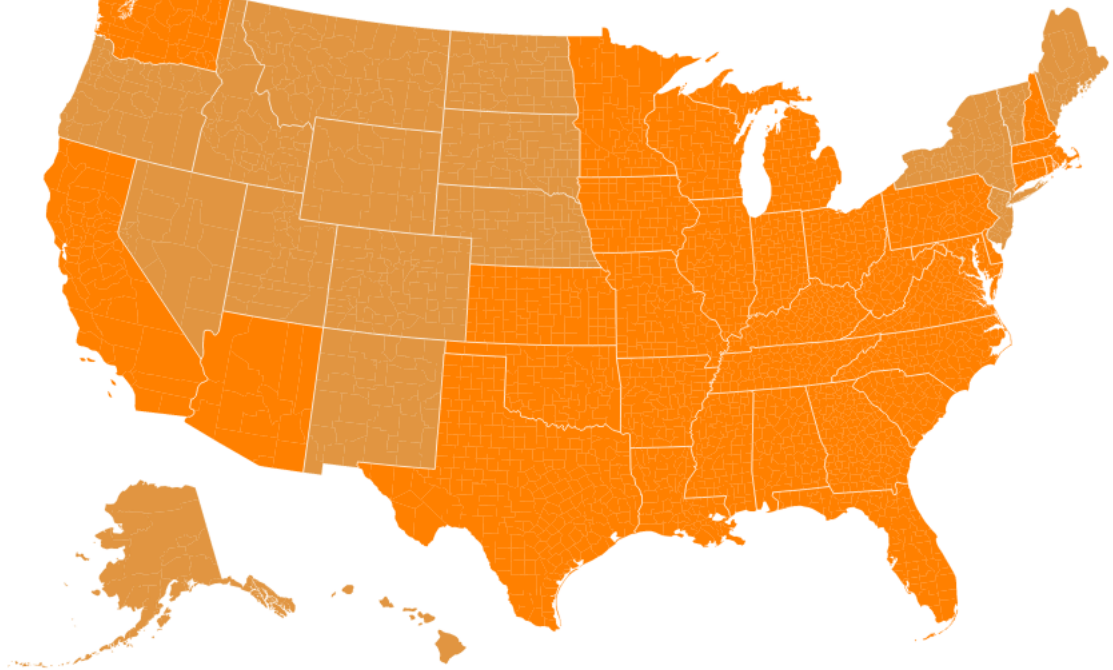
## United States

1,639,648 confirmed cases

F →

| | |
|---|---|
| Less than 25% Reduction in Average Mobility (Based on Distance Traveled) | F → |
| Less than 55% Reduction in Non-Essential Visits | F → |
| Less than 40% Decrease in Encounters Density Compared to National Baseline | F → |

Share This View




States Counties

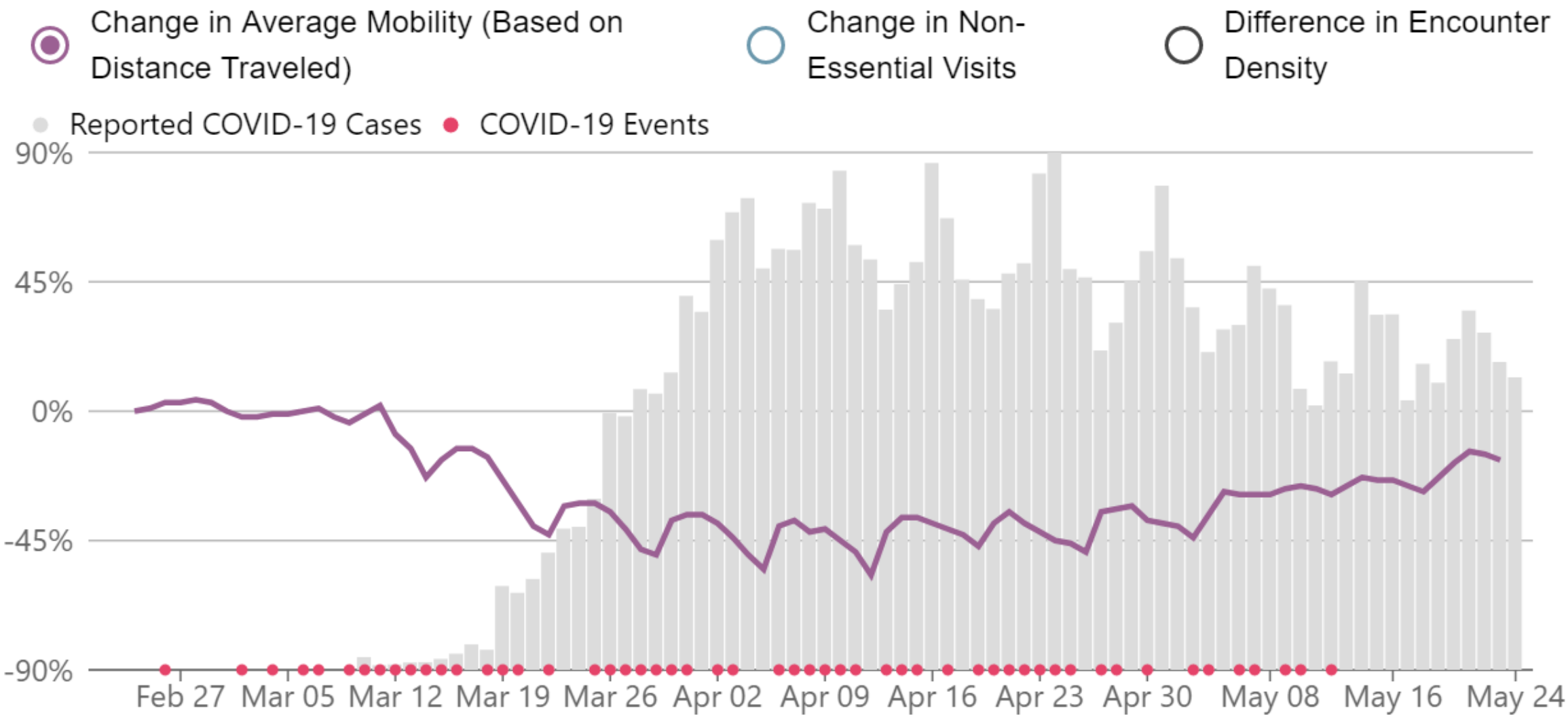
| | |
|---|---|
| District of Columbia | C- ↘ |
| Montana | D+ → |

◉ Change in Average Mobility (Based on Distance Traveled)

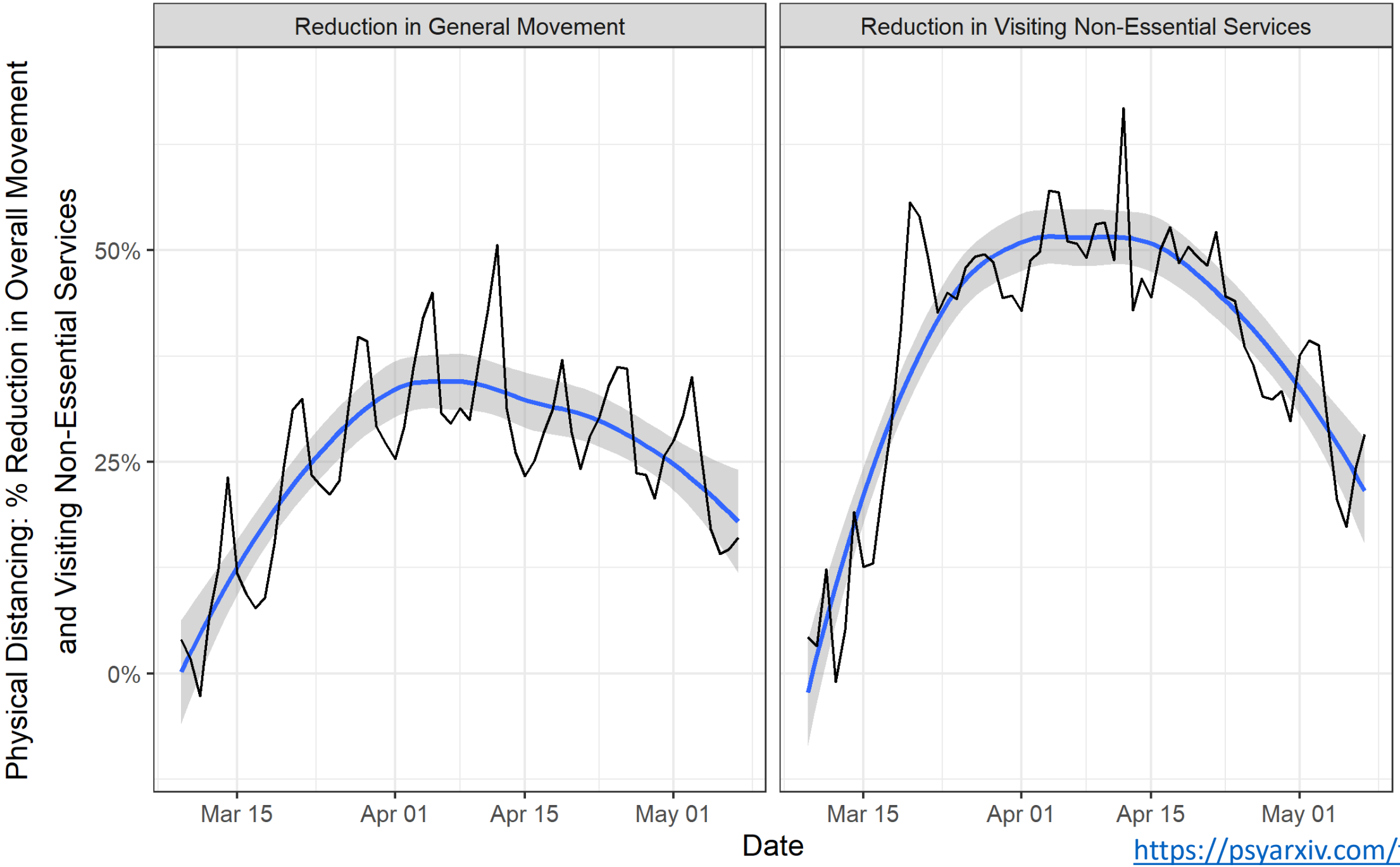
○ Change in Non-Essential Visits

○ Difference in Encounter Density

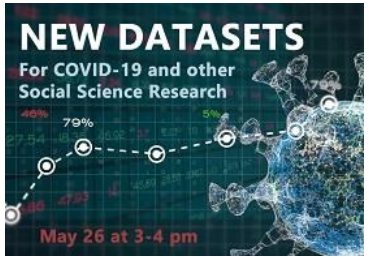
Unacast


Change in Average Mobility (Based on Distance Traveled)
Change in Non-Essential Visits
Difference in Encounter Density
Reported COVID-19 Cases
COVID-19 Events
90%
45%
0%
-45%
-90%
Feb 27
Mar 05
Mar 12
Mar 19
Mar 26
Apr 02
Apr 09
Apr 16
Apr 23
Apr 30
May 08
May 16
May 24

Unacast




https://psyarxiv.com/t3yxa/

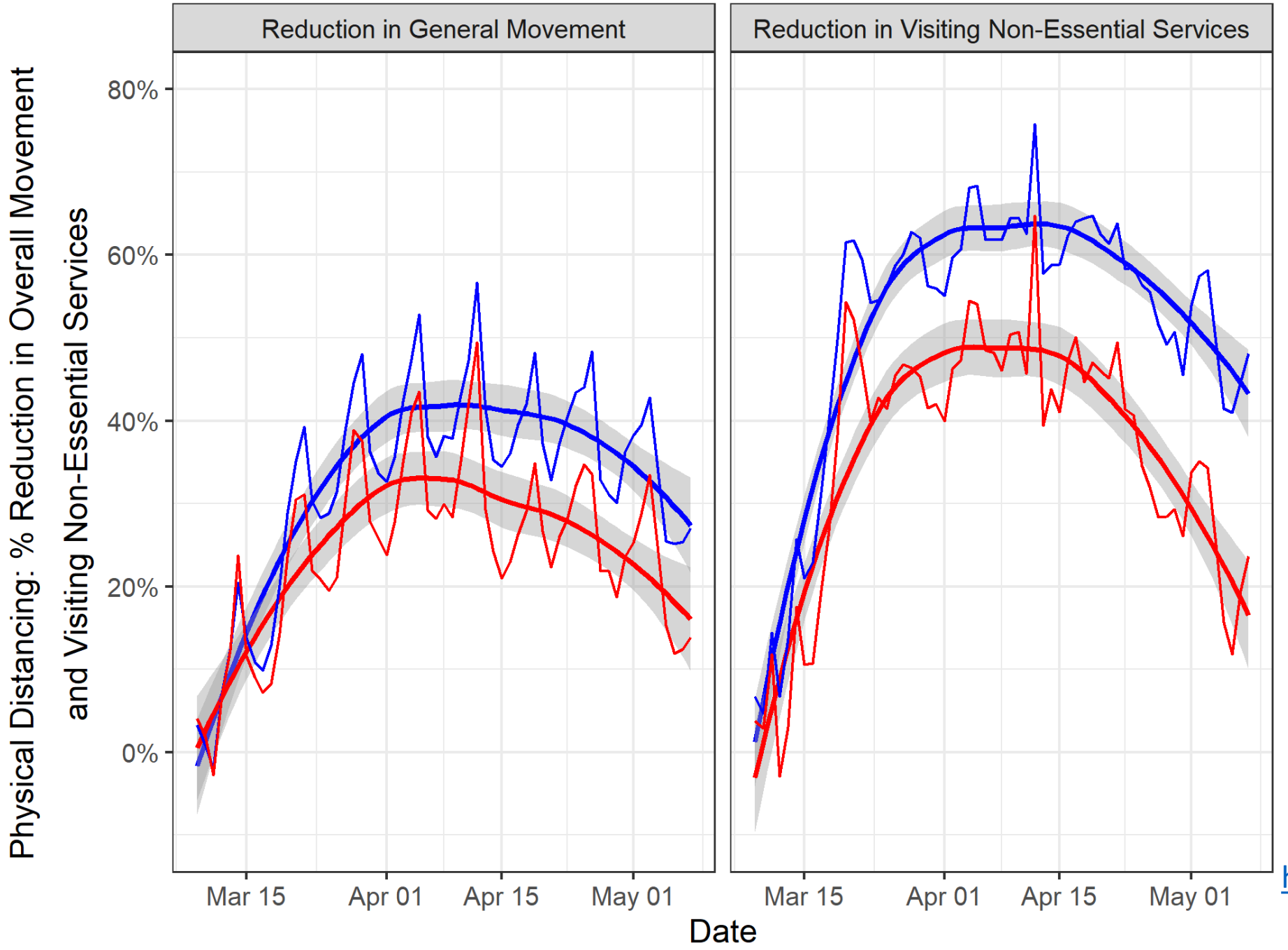


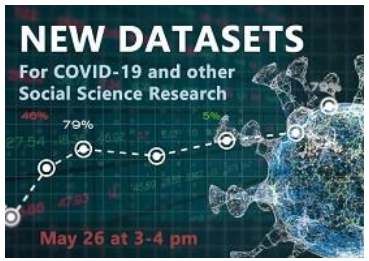

May 26, 2020

# Does partisanship predict social distancing in geo-location data?

https://psyarxiv.com/t3yxa/

**Anton Gollwitzer** Yale

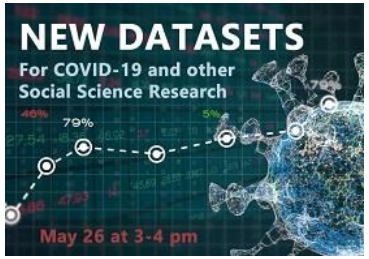



May 26, 2020

https://psyarxiv.com/t3yxa/

**Anton Gollwitzer** Yale

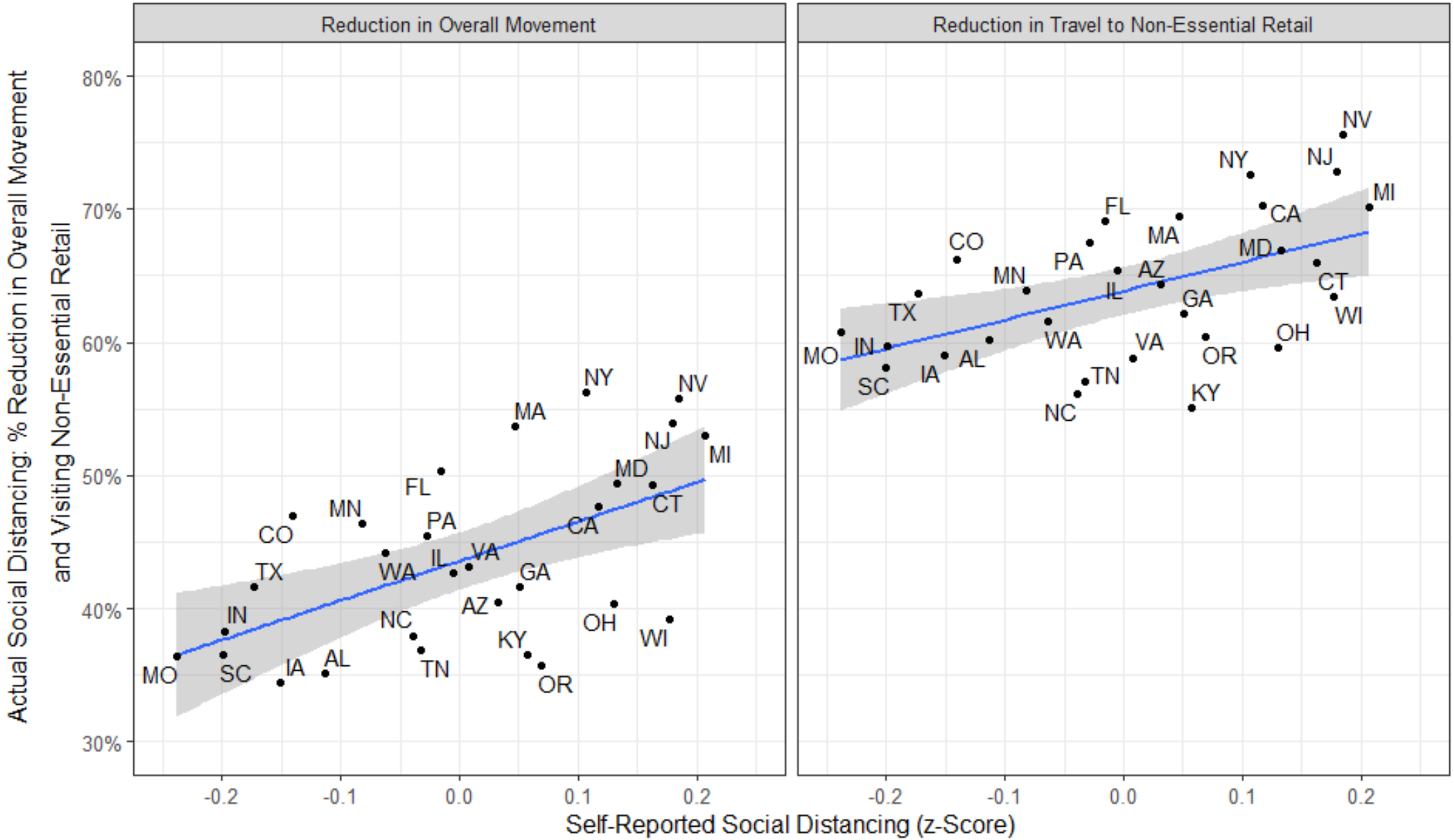



May 26, 2020

# Do geo-location data track self-reported social distancing?

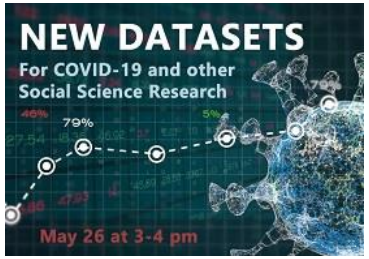
https://psyarxiv.com/kvnwp/

**Anton Gollwitzer** Yale




https://psyarxiv.com/kvnwp/

# Further Documentation/Uses




May 26, 2020

- Example article using geo-tracking data

- Example article using geo-tracking data

- Article calling for governments to use geo-location to combat COVID-19

- Article linking geo-tracking data to self-reported social distancing

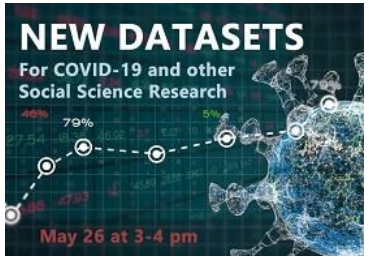
- General source for COVID-19 datasets

Contact: anton.gollwitzer@yale.edu

**Anton Gollwitzer** Yale

# SafeGraph.com: Location and Time-use Data from Smart Devices

https://www.safegraph.com/covid-19-data-consortium




May 26, 2020

Aggregated and anonymous data.
Unit of observation is the location.
https://docs.safegraph.com/docs/social-distancing-metrics
https://docs.safegraph.com/docs/places-schema

They have a number of products – Roy Lederman is working on Yale wide agreement.

Foot traffic product (that I have not used)

- https://github.com/SafeGraphInc/SafeGraphR

Core, Patterns, and Geometry products.
**Core** – business info on 5.9M locations & can be linked to the Yale licensed REFUSA data (started doing this).
**Geometry** – polygons for 5.9M locations (a bit more work to get and we have not used).
**Patterns** – (see column)
**Home dwell time products** (these have been changing a bit).

- Home & Not Home dwell time empirical distribution and median for CBG

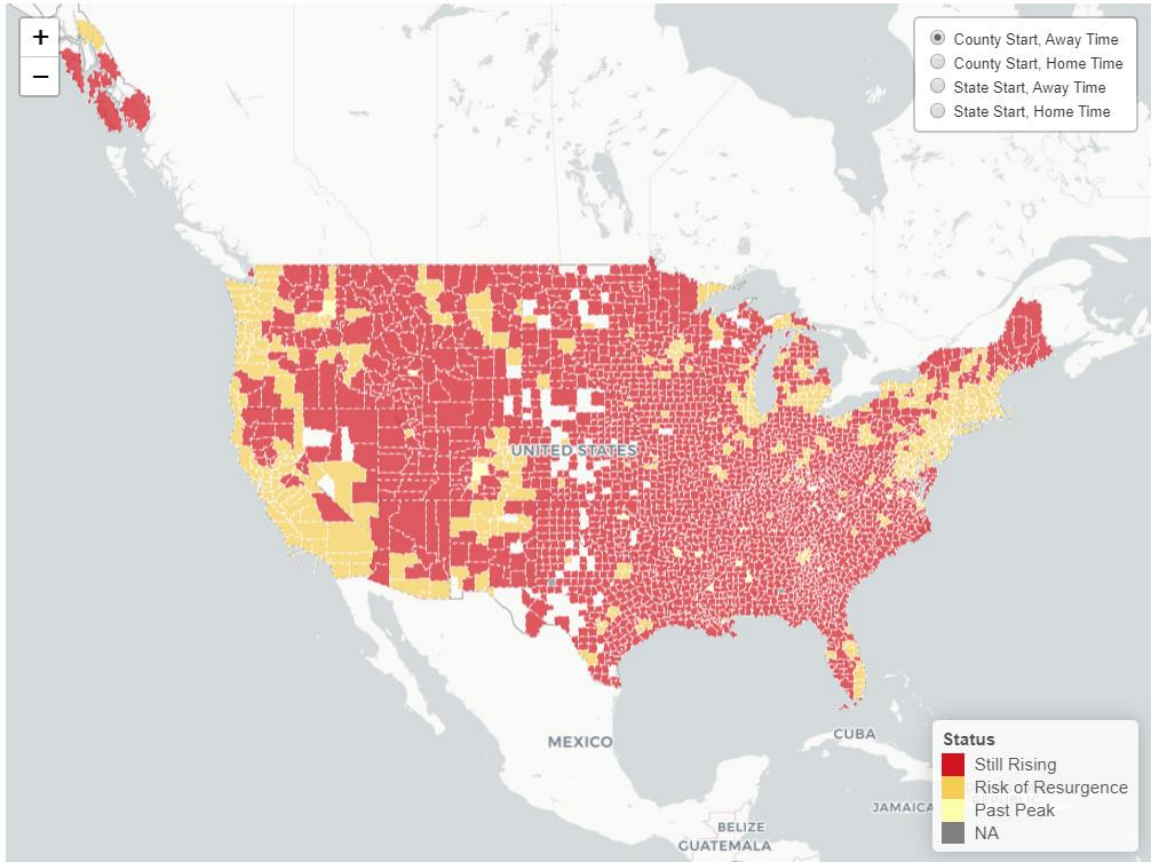
**Patterns**

- Place_id
- Location_name
- Street_address
- City
- Region
- Postal_code
- Brand info
- Dates for binning
- Visitor counts (not employees, but employees soon)
- Visits by day
- CBG (Census Block Group)
- Visitor home CBG
- Time of day by CBG
- Distance from home (median)
- Dwell time buckets
- Other places visited same day and month
- Popularity by hour
- Device info

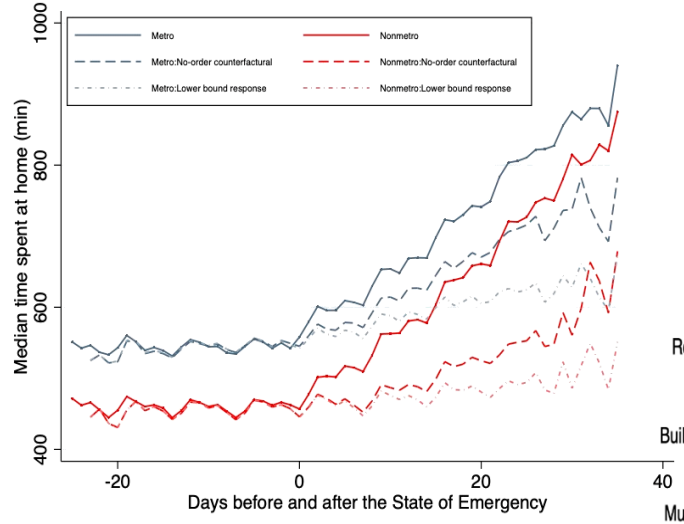
# Applications

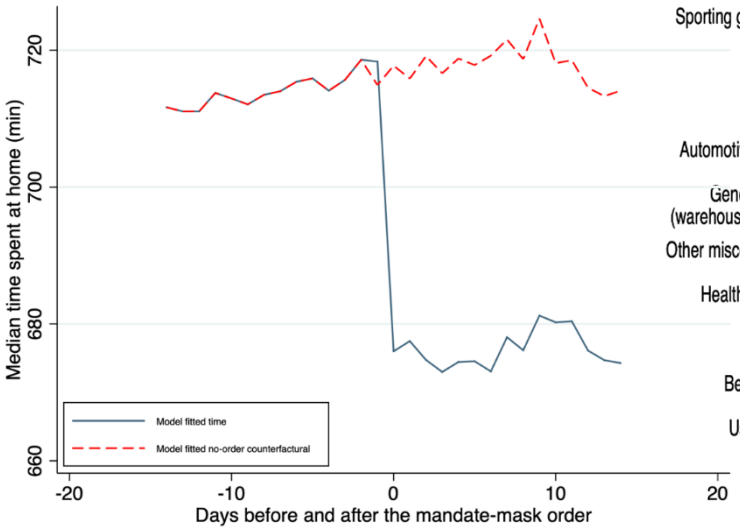
https://jbayham.github.io/maps/distancing/

Mobile Device COVID-19 Simulations by US Counties



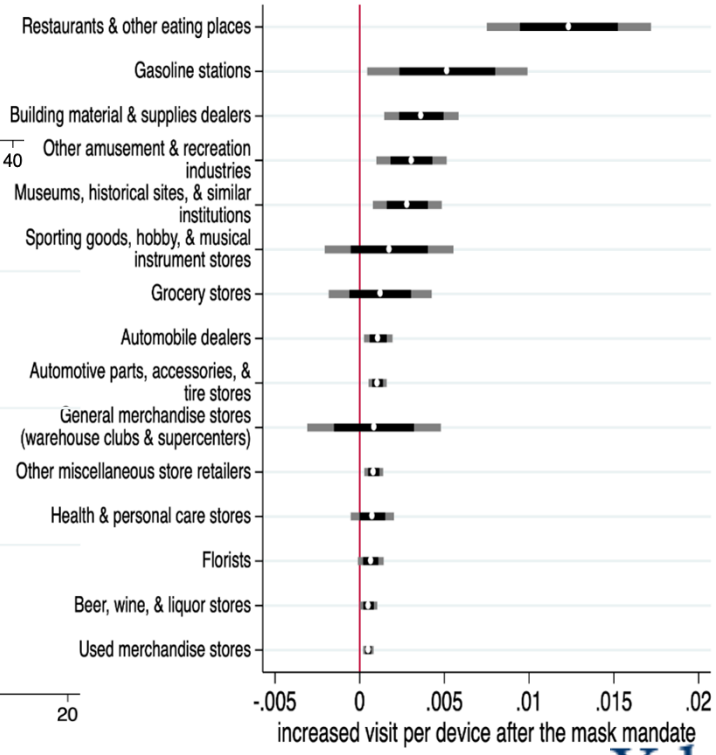

Fenichel, E.P., Berry, K., Bayham, J., Gonsalves, G., 2020. A cell phone data driven time use analysis of the COVID-19 epidemic. medRxiv https://www.medrxiv.org/content/10.1101/2020.04.20.20073098v1.




Yan, Y., Malik, A.A., Bayham, J., Fenichel, E.P., Couzens, C., Omer, S.B., 2020. Measuring voluntary social distancing behavior during the COVID-19 pandemic. medRxiv https://www.medrxiv.org/content/10.1101/2020.05.01.20087874v1.

Yan, Y., Bayham, J., Fenichel, E.P., Richter, A., 2020. Do Face Masks Create a False Sense of Security? A COVID-19 Dilemma. Medrxiv https://www.medrxiv.org/content/10.1101/2020.05.23.20111302v1.




**Eli Fenichel & Roy Lederman** Yale

# Other Ideas

In the works

- Examining compensatory behavior while opening up.
- Identifying locations for Mobile Testing Intercepts.
- Estimating demand for outdoor recreations and willingness to travel.
- Refine our Simulation Model.

Data have some biases

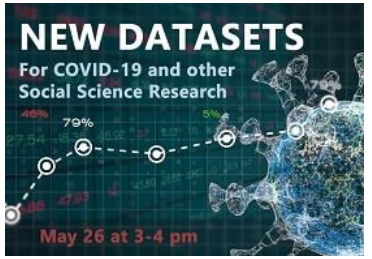
- https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#offline=true&sandboxMode=true
- Connecting to other data
  - Weather data processed (need to find the public version).
  - REFUSA
  - Case & Testing Data.
- Working to acquire other smart device data
  - UberMedia Close Contacts
  - XMode tracks.

Locations for Mobile Test Intercepts (Top 3 Visited Cites in New Haven 4/26 – 5/3).

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Yale New Haven Hospital | General Medical and Surgical Hospitals | General Medical and Surgical Hospitals | 20 York St Fl 2 | New Haven | 6510 | NA | Yale New Haven Hospital | 4/26/2020 4:00 | 5/3/2020 4:00 | 2 | 386 |
| 2 | Hammonasset Beach State Park | Museums, Historical Sites, and Similar Institutions | Nature Parks and Other Similar Institutions | Hammonasset Beach State Park | New Haven | 6443 | NA | NA | 4/26/2020 4:00 | 5/3/2020 4:00 | 6 | 274 |
| 3 | Bay River Marketplace | Lessors of Real Estate | Malls | 36439-26 Mile Road | New Haven | 48048 | NA | NA | 4/26/2020 4:00 | 5/3/2020 4:00 | 6 | 254 |

**Eli Fenichel & Roy Lederman** Yale

# Data, GIS and Statistical Support




May 26, 2020

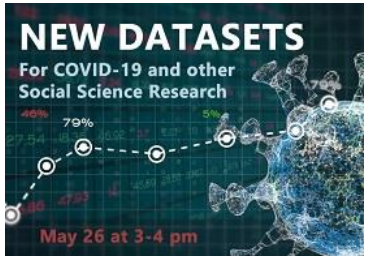
## Find, use and manage your research data

- Based out of Marx library (formerly CSSSI)
- YUL Working remotely – online services and support

Who to contact:

Barbara Esty, Data Librarian, barbara.esty@yale.edu
Miriam Olivares, GIS Librarian, gishelp@yale.edu
Sara Gottlieb-Cohen, Manager, Statistical Support, sara.gottlieb@yale.edu

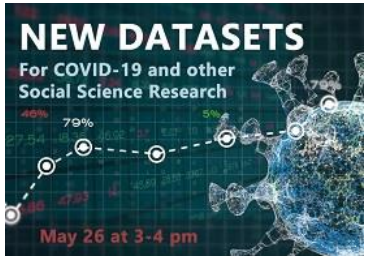
# Finding Data




May 26, 2020

- Quicksearch - Search the library catalog for data sets – Learn more.
- Research Guides:
  - Social Science Data - Sources of economic, elections, political, and other Social Science data.
  - U.S. Census - Access to statistics from the U.S. Census through multiple sources.
  - Covid-19 - working collection for researchers looking to explore the impact of COVID-19 across social science disciplines.
- Acquiring data – library purchases, navigating data use agreement process
- Accessing data – mediated access to data, reformatting

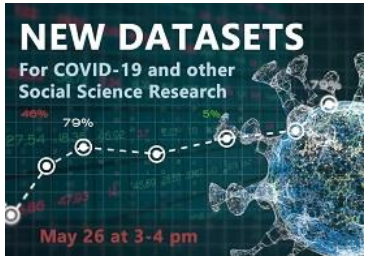
# Using Data




May 26, 2020

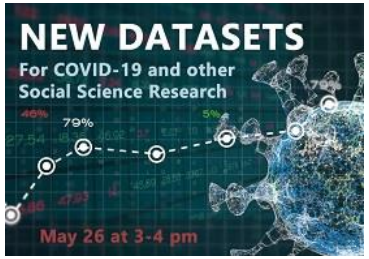
- Data and Subject librarians
  - Trouble shooting
  - Finding Supporting and additional data/documentation
- GIS at Yale
  - Software installation
  - Using GIS resources
- Statistical Support
  - Statistical consultants
  - Workshops

# Managing Data




May 26, 2020

- Research Data Management - Find tips and resources for managing your research data, wherever it came from.
  - DMPTool - templates and guidance for Data Management Plan (DMP)
  - Repositories
    - Dryad – Yale is an institutional member
    - Other suggestions based on need



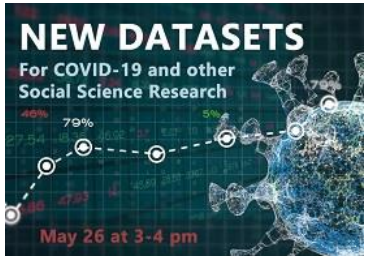

May 26, 2020

# Statistical Support Services at Marx Library

sara.gottlieb@yale.edu

**Sara Gottlieb-Cohen** Yale

# Drop-in consultation sessions




May 26, 2020

Home » Data, GIS & Statistical Support » Statistical Support » Walk-In Help » Walk-In Help Schedule

- Statistical Support
  - About Us
  - Workshops
  - Walk-In Help
    - Get Ready
    - Walk-In Help Schedule
- GIS at Yale
- Research Data Management
- Science Data
- Social Science Data

## Walk-In Help Schedule

**UPDATE (3/10/20): All consultations will be held VIRTUALLY for the foreseeable future using Microsoft Teams (these sessions are noted in purple). To join a virtual session, please follow the instructions provided in each calendar event.**

Walk-in help is available at Marx Library, Rosenkranz Hall, and the Cushing/Whitney Medical Library. These sessions are free to anyone affiliated with Yale University. Walk-in sessions are usually limited to 20-30 minutes so other patrons can be served. The current walk-in schedule can be found below:

Campus | Category | Audience

< > Today **May 2020** Agenda | Day | Week | Month

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|---|---|---|---|---|---|---|
| 26 | 27<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 28<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 29<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 30<br>8am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>6pm VIRTUAL StatLab | 1<br>9am VIRTUAL StatLab<br>12:30pm VIRTUAL Sta | 2 |
| 3 | 4<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 5<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 6<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 7<br>8am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>6pm VIRTUAL StatLab | 8<br>9am VIRTUAL StatLab<br>12:30pm VIRTUAL Sta | 9 |
| 10 | 11<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 12<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 13<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 14<br>8am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>6pm VIRTUAL StatLab | 15<br>9am VIRTUAL StatLab<br>12:30pm VIRTUAL Sta | 16 |
| 17 | 18<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 19<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 20<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 21<br>8am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>6pm VIRTUAL StatLab | 22<br>9am VIRTUAL StatLab<br>12:30pm VIRTUAL Sta | 23 |
| 24 | 25 | 26<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 27<br>9am VIRTUAL StatLab<br>1pm VIRTUAL StatLab | 28<br>8am VIRTUAL StatLab<br>1pm VIRTUAL StatLab<br>6pm VIRTUAL StatLab | 29<br>12:30pm VIRTUAL Sta | 30 |
| 31 | 1 | 2 | 3 | 4 | 5 | 6 |

https://marx.library.yale.edu/statistical-consulting-services

## VIRTUAL StatLab consultation hours: Sara Gottlieb-Cohen

VIRTUAL StatLab consultation hours: Sara Gottlieb-Cohen

Sara Gottlieb-Cohen will be available virtually for StatLab consultations.

Please follow these steps to access the StatLab on Microsoft Teams:

- Visit http://bit.ly/StatLab-Team in your web browser; you may also follow the instructions to dowload the desktop app
- Log in using your Yale email and password
- On the landing page, type a message to alert the consultant on duty that you have arrived and would like help
- The consultant will start a video chat session with you, and will likely ask that you share your screen. Be prepared that you may have to wait if the consultant is currently assisting another patron.

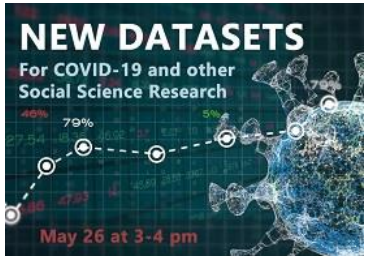
If you have any issues, please call Sara Gottlieb-Cohen at (203) 432-3278.

**Date:** Monday, May 4, 2020 Show more dates
**Time:** 9:00am - 1:00pm
**Campus:** Science Hill

Browse/Search for more events

**Sara Gottlieb-Cohen** Yale

# Workshops




May 26, 2020

## R

- First steps with R
- Second steps with R
- Hypothesis testing with R
- Data manipulation using the tidyverse
- Data visualization using ggplot2

## Python

- First steps with Python
- Second steps with Python
- Python for data science
- Web scraping using Python

## Stata

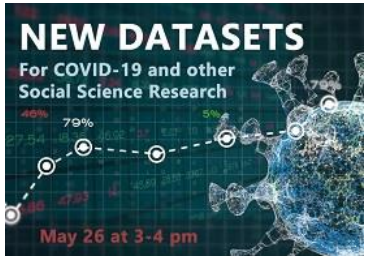
- First steps with Stata
- Second steps with Stata

## Other

- Survey design & Qualtrics
- Qualitative data analysis using NVivo

**Sara Gottlieb-Cohen** Yale

# Summer series:
# Data analysis using R




May 26, 2020

A weekly event geared toward people who have some experience with R but want additional practice on how to apply coding skills to answer different research questions.

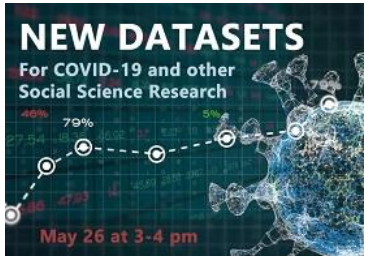
The specific topics, tests, and data techniques will vary weekly, but expect to cover:

- Deciding on a statistical test that will appropriately answer a research question
- Selecting appropriate packages
- Manipulating data
- Conducting and interpreting descriptive and inferential statistics
- Visualizing data

https://schedule.yale.edu/event/6720405

**Sara Gottlieb-Cohen** Yale

# New Datasets for COVID-19 and other Social Science Research




May 26, 2020

## Thank you!

All session information will be posted here:
https://isps.yale.edu/new-datasets

To indicate your interest in future research opportunities, please go to:
https://isps.yale.edu/social-science-research-opportunities

Yale

Questions? Contact limor.peer@yale.edu