China Data Institute

Center for Geographic Analysis
Harvard University

社会地理计算联合研究中心
Geocomputation Center for Social Science

Foursquare Community Mobility Data In COVID-19 Research

Xiaoyue Wang, Wuhan University, SooyeWang@whu.edu.cn
the Mobility Research Group of the ChinaDataLab
the Geo-computation Center for Social Sciences, Wuhan University

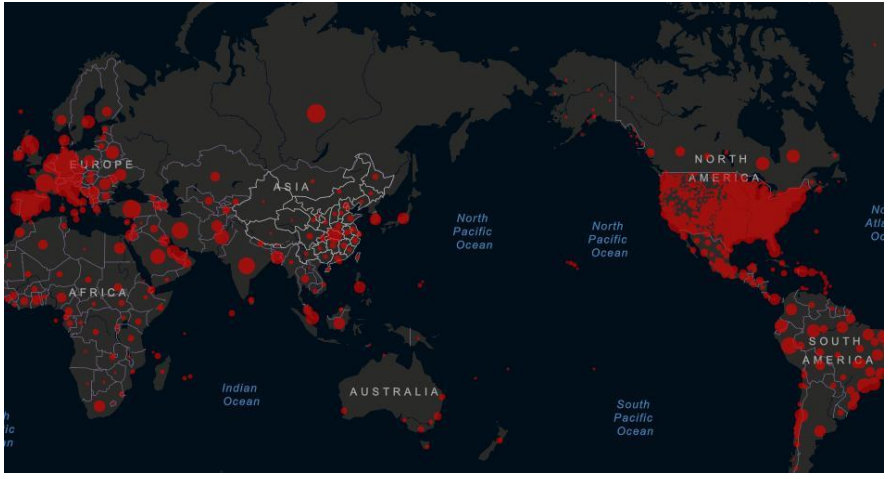
04-17-2021（GMT+8）

Outline

- ❑ Background
- ❑ Data Sources
- ❑ Data Depository
- ❑ Related research
- ❑ Data Workflows
- ❑ References

Background

The COVID-19 pandemic poses unprecedented challenges for countries around the world.
Many studies have shown that mobility data can provide significant supports for public health actions across the pandemic. Users' locations are usually collected from public transit, mobile operator, and mobile phone applications.



Global Cases
127,289,043

Cases by Country/Region/Sovereignty	
30,263,027	US
12,534,688	Brazil
12,039,644	India
4,606,196	France
4,477,916	Russia
4,347,014	United Kingdom
3,532,057	Italy
3,255,324	Spain
3,208,173	Turkey
2,789,480	Germany
2,382,730	Colombia
2,308,597	Argentina
2,267,964	Poland
2,226,550	Mexico
1,864,984	Iran
1,703,036	Ukraine

Background

In the COVID-19 studies, there are four common ways to estimate human Mobilities：

1） including public transit-based transportation flow (e.g., IATA)；

2） GPS-based social activities (e.g., Google Mobility Report) ；

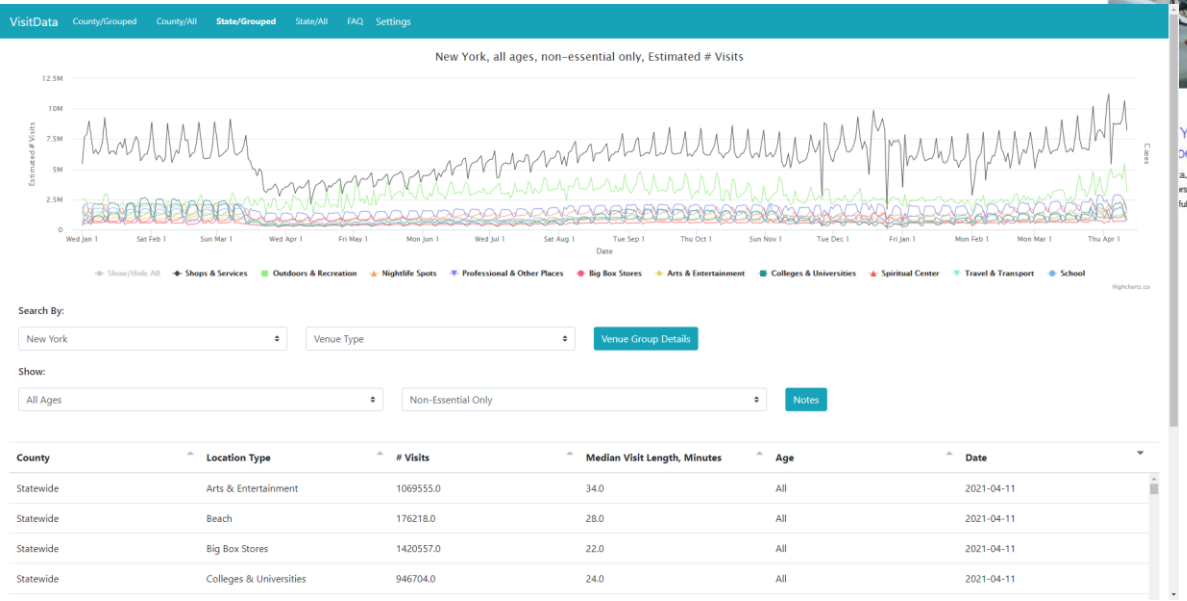
3） mobility index (e.g., Baidu Mobility Index)；

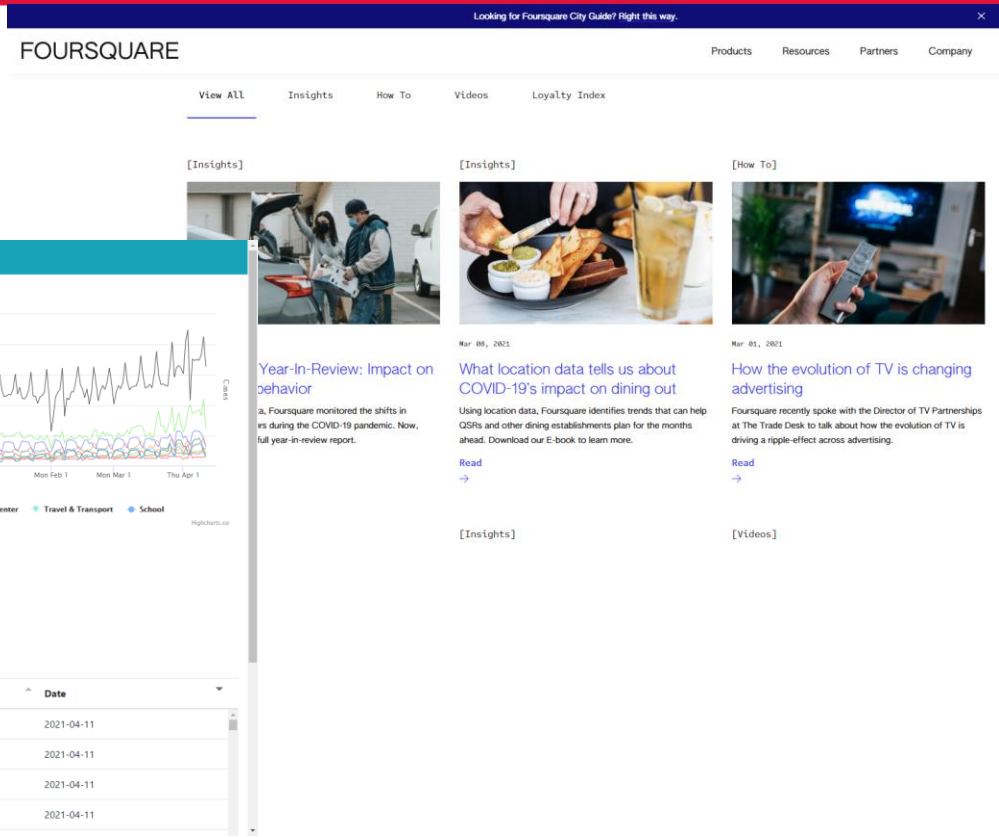
4） social media derived mobility index (e.g., Geotagged Tweets).

Category	Dataset name and provider	Region and scale	OD flow	Variables	Availability	Pros	Cons
Social activity	Apple Mobility Trends Report	63 countries	N	Transport frequency measures in three types of modes	Free	global wide; one single file; data divided by country/region, sub-region, city	data source method (requests for directions in Apple Maps)
	Google Mobility Reports	131 countries and regions (after Feb 15, 2020)	N	POI access frequency in six categories of the destinations	Free	global wide ； one single file	not comparable among countries
	Foursquare Mobility Reports	US at national level plus SF, NYC, LA, and Seattle	N	POI access frequency to 25 categories of destinations	Submit Application	Available in 25 types of POI and by age group	Only available in US
	SafeGraph Mobility Reports	US by state and county	Y	Foot traffic to businesses and consumer points of interest	Submit Application	Varieties of data categories	publicly accessible (application and Amazon S3)

Data Sources

- Foursquare Mobility Data
- https://visitdata.org/index.html
- https://foursquare.com/reports/





Data Sources

- **Foursquare Mobility Data：**
- **Region and scale：** US at national level plus SF, NYC, LA, and Seattle (state/county)
- **Update frequency：** Daily
- **Variables:** POI access frequency to 25 categories of destinations
- **Availability:** Submit Application
- **Location type：** Airport, Alcohol, Banks, Bus, Food, Gas Stations, Hotel, School……
- **Pros:** Available in 25 types of POI and by age group
- **Cons:** Only available in US and no OD flow

190+

Countries and 50 territories

900+

Venue Categories

2.4M

POI updates monthly, from ground truth sources

Data Sources

- Data Type:

County_avgDuration_XX

County_p50Duration_XX

County_visits_XX

State_avgDuration_XX

State_ p50Duration_XX

State_ visits_XX

(XX: POI type, eg: Airport, Alcohol, Banks, Bus, Food, Gas Stations, Hotel, School……)

NAME	2020/1/3	2020/1/4	2020/1/7	2020/1/8	2020/1/9	2020/1/10	2020/1/11	2020/1/12
Autauga C	8253	9721	8316	8263	8598	9196	8021	8465
Baldwin Co	124988	139532	109660	101448	96337	116829	104916	128188
Barbour County								
Bibb County								
Blount Cou	17739	11062	10951	8001	11655	14205	15357	16324
Bullock County								
Butler County								
Calhoun Co	45992	56383	49063	43104	44259	49521	53470	51486
Chambers	8648	6905			6693	6977		
Cherokee County								
Chilton Cou	16489	15286	13688	13063	14477	15654	15911	12584

Data Sources

- Data Type(by age group):

  18_24

  24_34

  35_44

  45_54

  55_64

  65_plus

state	age	dt	norm_visits
WY	45_54	6/15/2020	265683.9934
WY	45_54	10/7/2020	291684.4208
WY	35_44	10/14/2020	207314.1895
WY	35_44	3/4/2020	198642.9371
WY	35_44	2/8/2020	195769.9286
WY	25_34	11/11/2020	222534.539
WY	35_44	10/31/2020	205468.786
WY	18_24	2/11/2020	173655.6054
WY	18_24	3/26/2020	107729.2758
WY	55_64	11/16/2020	184468.6846
WY	25_34	10/8/2020	257624.0369
WY	45_54	10/3/2020	310217.8606
WY	25_34	8/26/2020	266577.3784
WY	45_54	10/25/2020	170812.3829
WY	55_64	11/12/2020	194615.8962
WY	25_34	2/15/2020	236854.7695
WY	35_44	5/13/2020	162810.2325
WY	35_44	9/14/2020	248193.3952
WY	55_64	10/15/2020	257024.923
WY	55_64	9/8/2020	260614.2338
WY	25_34	6/3/2020	250791.9702
WY	25_34	8/7/2020	314046.9811
WY	65_plus	5/9/2020	148980.4395
WY	35_44	4/5/2020	137866.5919
WY	65_plus	2/19/2020	239567.0672
WY	55_64	10/30/2020	219844.252
WY	65_plus	9/28/2020	267822.1773
WY	65_plus	9/23/2020	293890.3347
WY	65_plus	7/22/2020	344676.1198
WY	25_34	5/8/2020	212143.5236
WY	55_64	9/22/2020	295725.2272

Data Depository: dataverse.harvard.edu

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts.* A personal dataverse is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal dataverse?

Journals

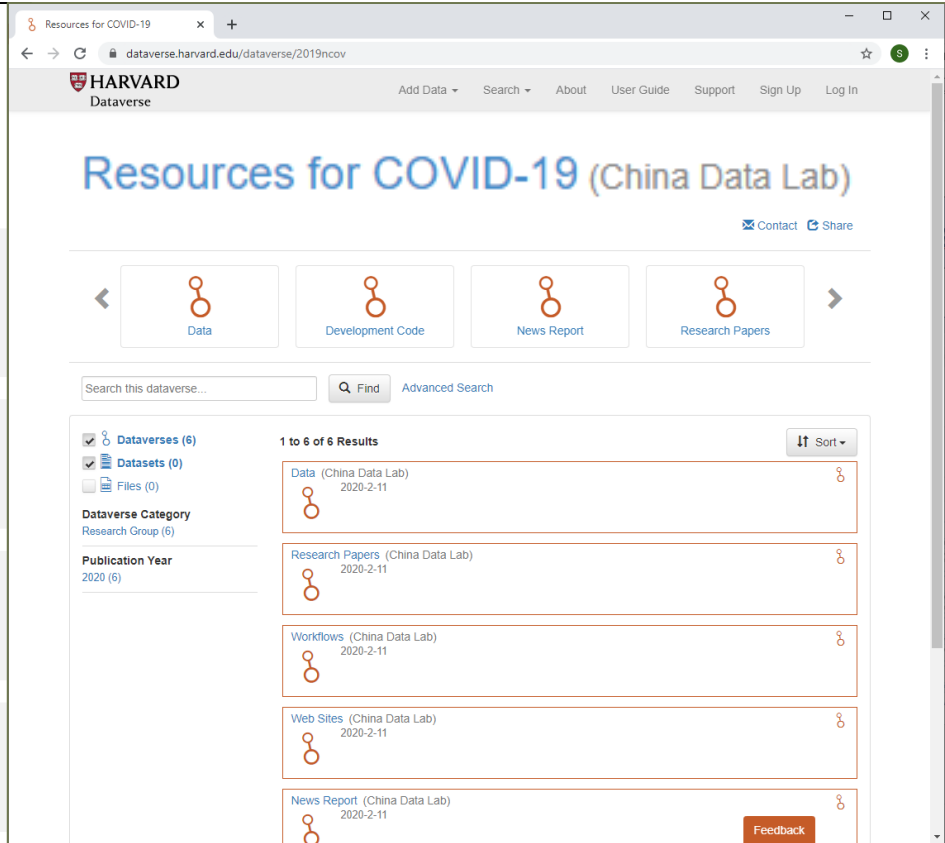
Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using Dataverse as part of your journal data policy or list of repository recommendations. Want to find out more about journal dataverses?

Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with Dataverse. Want to contribute?



Related research

- Age groups that sustain resurging COVID-19 epidemics in the United States

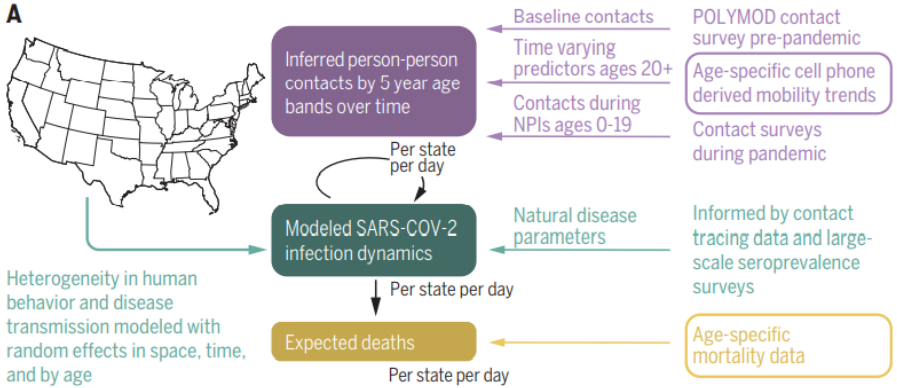
M. Monod et al., Science 371, eabe8372 (2021). DOI: 10.1126/science.abe8372

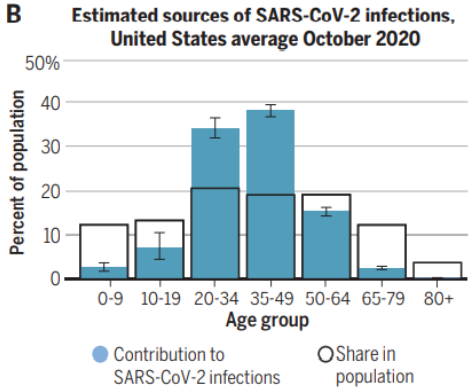
RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Age groups that sustain resurging COVID-19 epidemics in the United States

Mélodie Monod*, Alexandra Blenkinsop*, Xiaoyue Xi*, Daniel Hebert*, Sivan Bershan*, Simon Tie
Marc Baguelin, Valerie C. Bradley, Yu Chen, Helen Coupland, Sarah Filippi, Jonathan Ish-Horowicz
Martin McManus, Thomas Mellan, Axel Gandy, Michael Hutchinson, H. Juliette T. Unwin,
Sabine L. van Elsland. Michaela A. C. Vollmer, Sebastian Weber, Harrison Zhu, Anne Bezancon,
nan†, Samir Bhatt†,
College COVID-19 Response Team





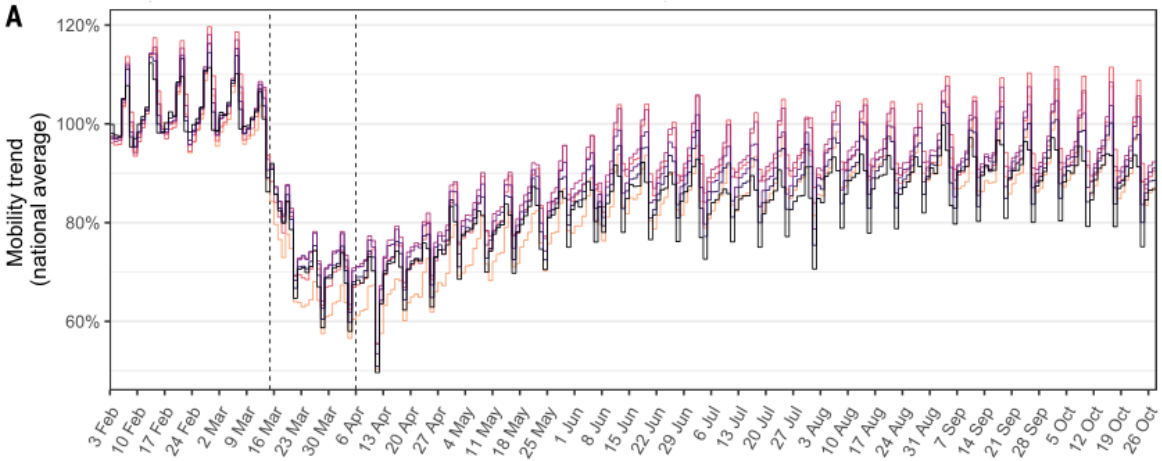
Model developed to estimate the contribution of age groups to resurgent COVID-19 epidemics in the United States. (**A**) Model overview. (**B**) Estimated contribution of age groups to SARS-CoV-2 transmission in October.

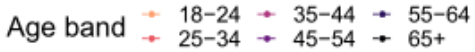
Related research

- INTRODUCTION: After initial declines, in mid-2020, a sustained resurgence in the transmission of novel coronavirus disease (COVID-19) occurred in the United States. Throughout the US epidemic, considerable heterogeneity existed among states, both in terms of overall mortality and infection, but also in the types and stringency of nonpharmaceutical interventions. Despite these stark differences among states, little is known about the relationship between interventions, contact patterns, and infections, or how this varies by age and demographics. A useful tool for studying these dynamics is individual, age-specific mobility data. In this study, we use detailed mobile-phone data from more than 10 million individuals and establish a mechanistic relationship between individual contact patterns and COVID-19 mortality data.

Related research

- Data:
- 1. Age-specific mortality data.
- 2. Age-specific mobility data from Foursquare Inc.
- 3. Age-specific mobility data from Emodo Inc.





Related research

- Methods:

1. Statistical analysis of human mobility data and COVID-19–attributable death data
2. Contact-and-infection model
3. Bayesian inference
4. Generated quantities
5. Validation and sensitivity analyses

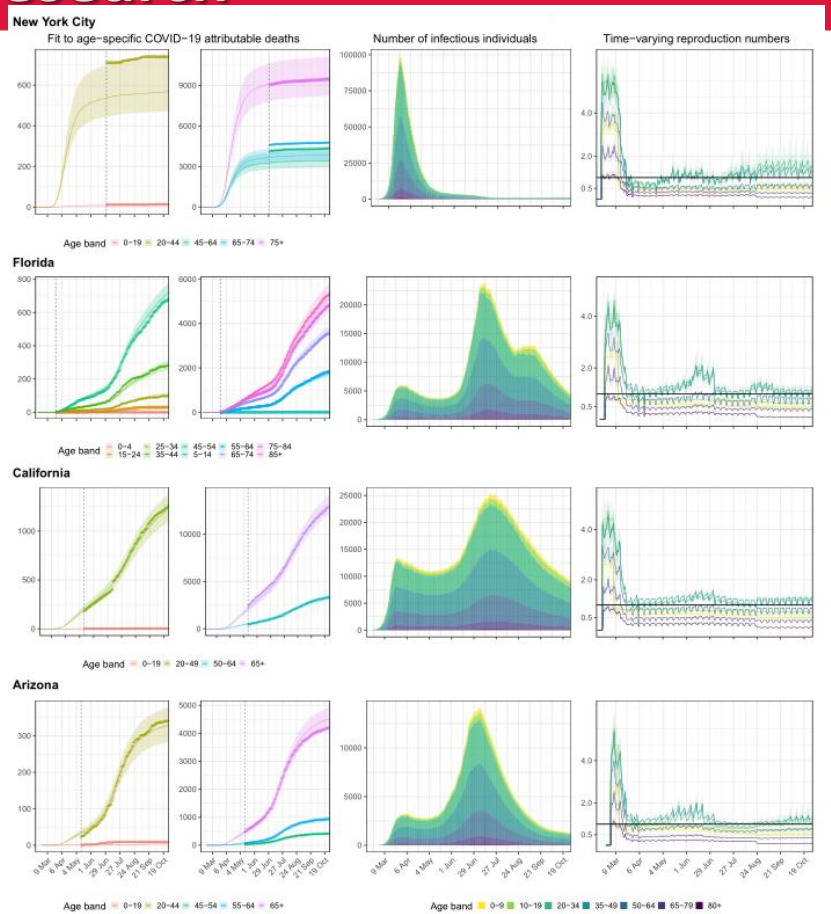


Fig. Model fits and key generated quantities for New York City, California, Florida, and Arizona.

Related research

- Result: We find that as of 29 October 2020, adults aged 20 to 34 and 35 to 49 are the only age groups that have sustained SARS-CoV-2 transmission with reproduction numbers (transmission rates) consistently above one. The high reproduction numbers from adults are linked both to rebounding mobility over the summer and elevated transmission risks per venue visit among adults aged 20 to 49. Before school reopening, we estimate that 75 of 100 COVID-19 infections originated from adults aged 20 to 49, and the share of young adults aged 20 to 34 among COVID-19 infections was highly variable geographically.

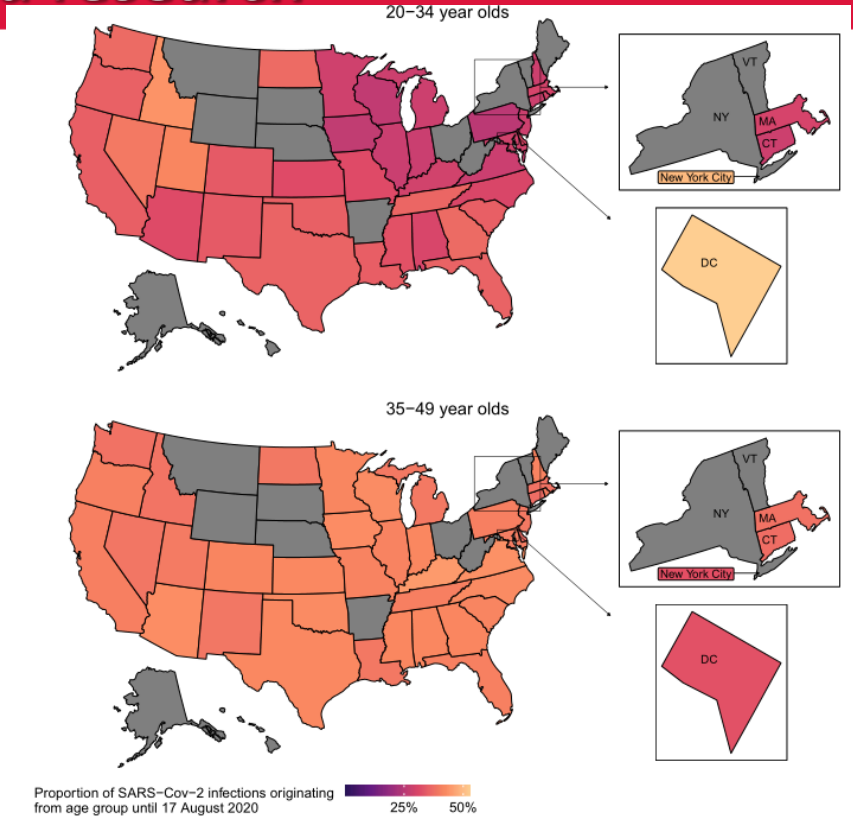
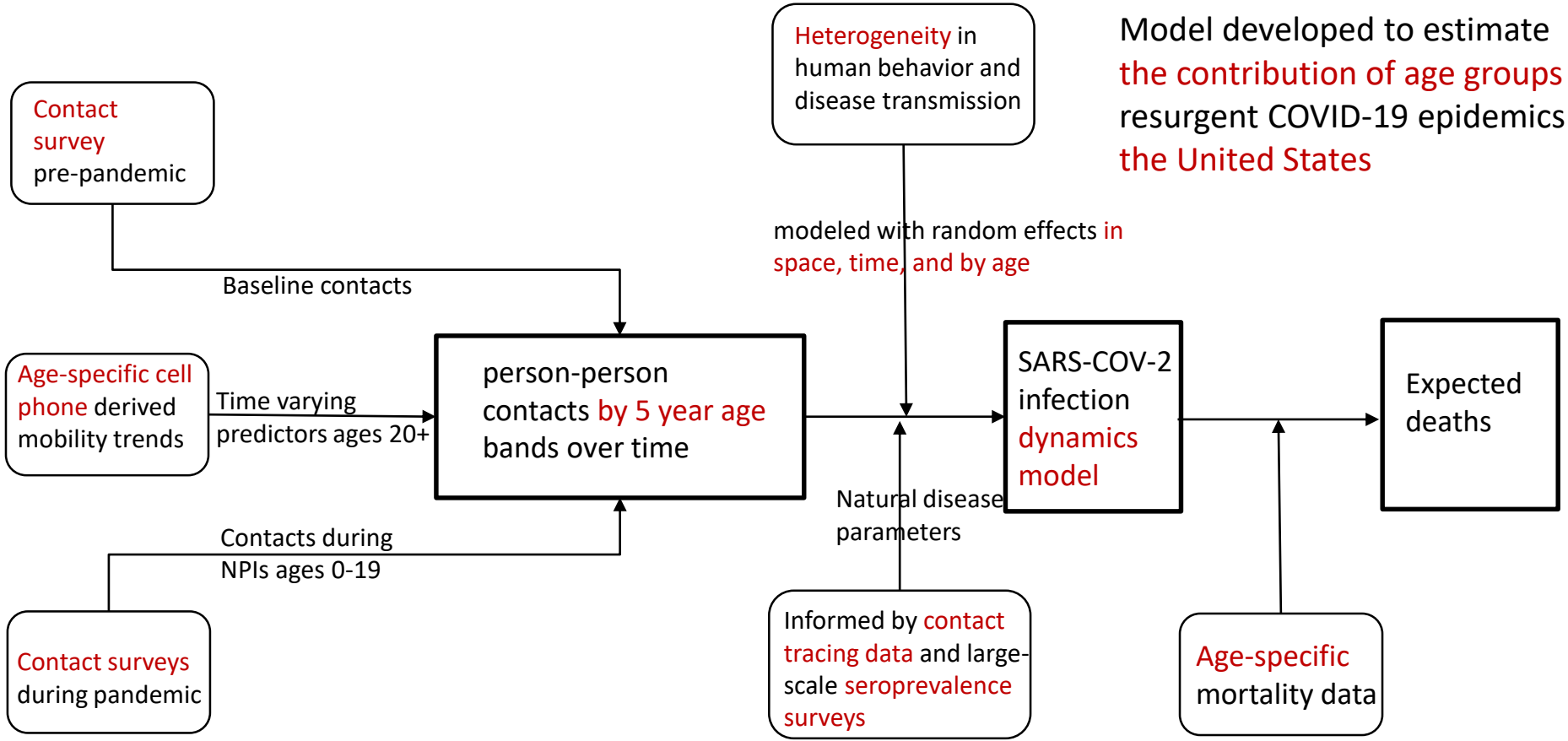


Fig. 4. Estimated spatial variation in the share of young adults aged 20 to 34 and adults aged 35 to 49 to COVID-19 spread until mid-August 2020. Posterior median estimates of the contribution to cumulated SARS-CoV-2 infections until 17 August 2020, prior to school reopening. State-level COVID-19 epidemics not considered in this study are in gray.

The Flowchart for Data Integration & Visualization

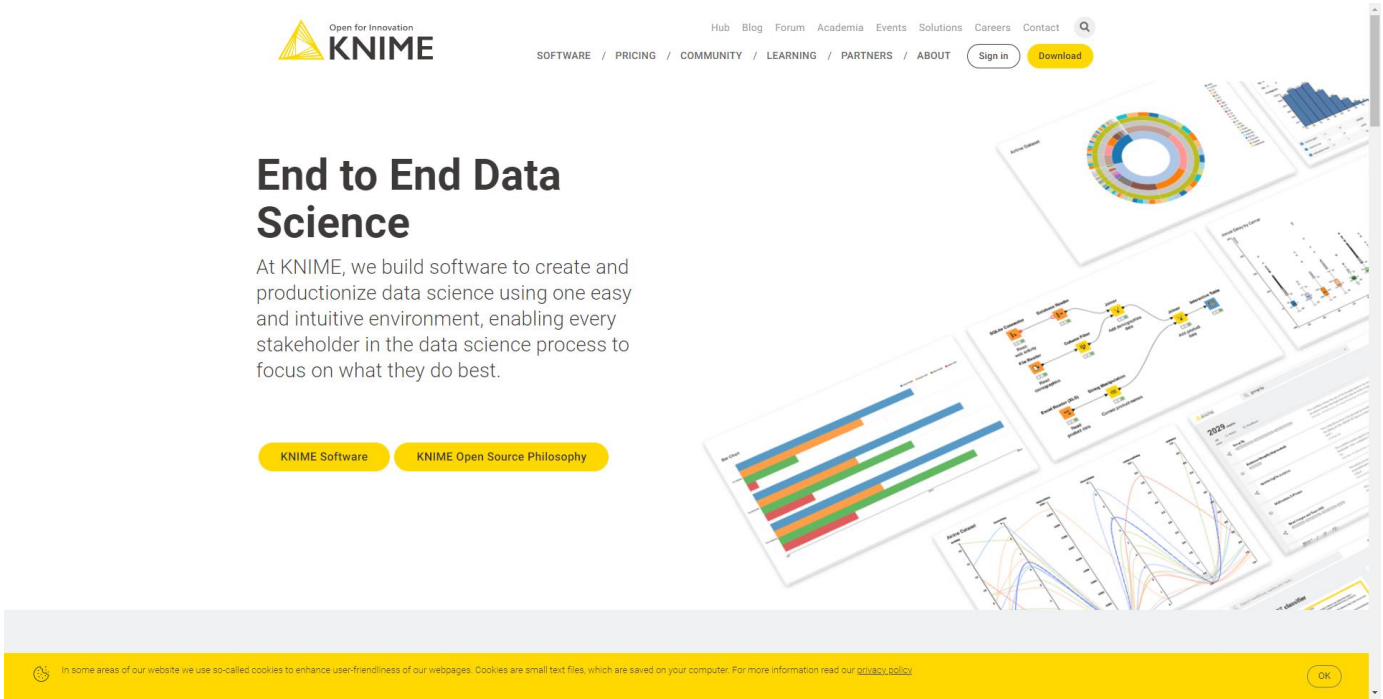


Model developed to estimate the contribution of age groups to resurgent COVID-19 epidemics in the United States

The Flowchart for Data Integration & Visualization

KNIME

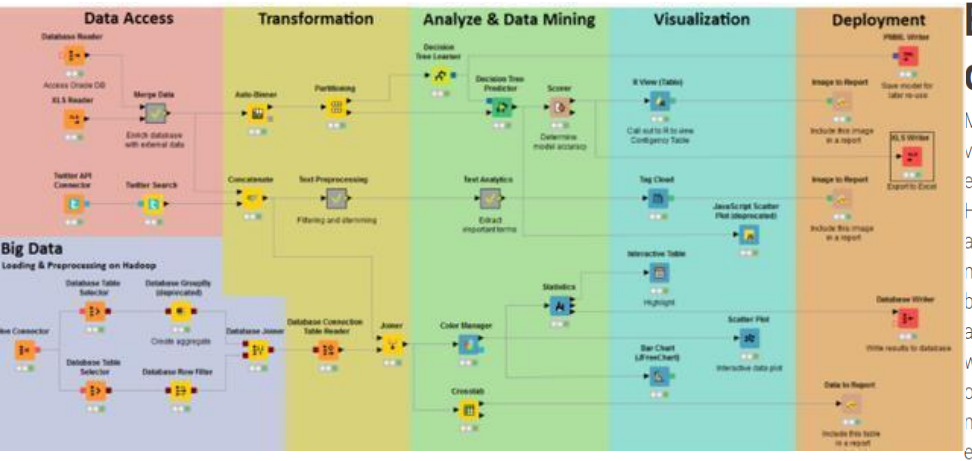
https://www.knime.com/



The Flowchart for Data Integration & Visualization

KNIME | Open for innovation

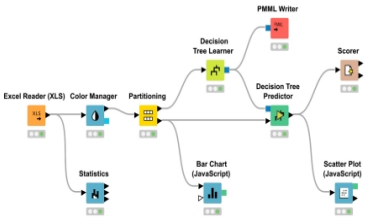
1.The freedom to use the tools you like
2.Retain quality and accuracy in your analytics
3.Spend your time where it makes the most impact
4.Easily deploy and scale your work



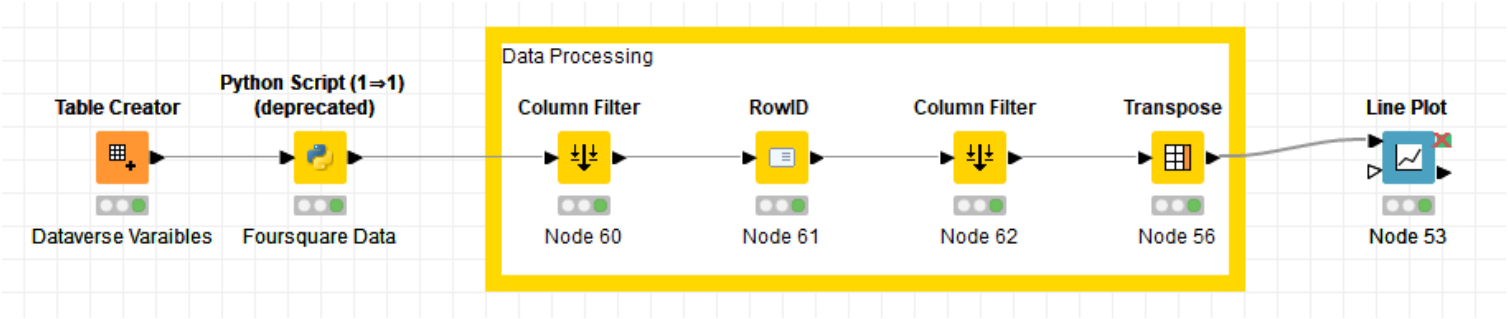
Analytics Expert

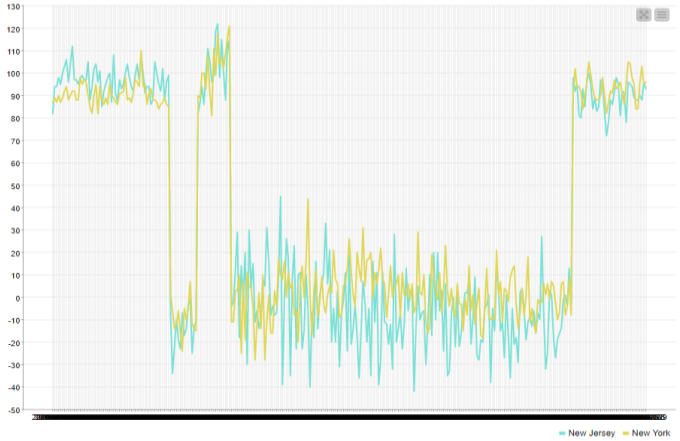
Extract insights from your data

Making sense of data requires sophisticated modeling and visualization techniques. Modern analytics now also encompasses machine learning and artificial intelligence. However, to solve certain problems, you still want to be able to reach out to the classic statistical analysis and data mining methods. That's why we continually add both bleeding edge algorithms and well established analytics and visualization methods to KNIME Software. Together with our R and Python integrations, plus integrations with other large, open source projects, you have the freedom to mix and match the tools you like, within one uniform environment.



Foursquare Data Access and Analysis with KNIME





Step 1: **Download KNIME 4.3.2 from** https://www.knime.com/downloads/download-knime
Step 2: **Download workflow (01_foursquare_data.knwf)** from CDL Harvard Dataverse or Google Drive
Step 3: **Open KNIME** from local PC or China Data Lab Cloud Platform
Step 4: **Import KNIME workflow** file (**01_foursquare_data.knwf**)
Step 5: **Configure "Input Data"** for each
Step 6: Click **Run** function from the top menu
Step 7: **Display the outputs**:

- **Table view** with state level average duration in airport over time

References

Hu, Tao, et al. "Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges." *ResearchGate Preprint, DOI* 10 (2021).

Li, Zhenlong, al. "ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility." *arXiv preprint arXiv:2104.05040* (2021).

Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., ... & Bao, S. (2021). Human mobility and COVID-19 transmission: a systematic review and future directions. *medRxiv*.

M. Monod et al., Science 371, eabe8372 (2021). DOI: 10.1126/science.abe8372

J. Mossong et al., Social contacts and mixing patterns relevant to the spread of infectious diseases. PLOS Med. 5, e74 (2008). doi: 10.1371/journal.pmed.0050074; pmid: 18366252.

J. Zhang et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science 368, 1481–1486 (2020). doi: 10.1126/science.abb8001;pmid: 32350060.