Measurement of Partisan Segregation of 180 million U.S. voters using advanced GIS Data Science

Devika Kakkar
Harvard University, CGA
kakkar@fas.harvard.edu

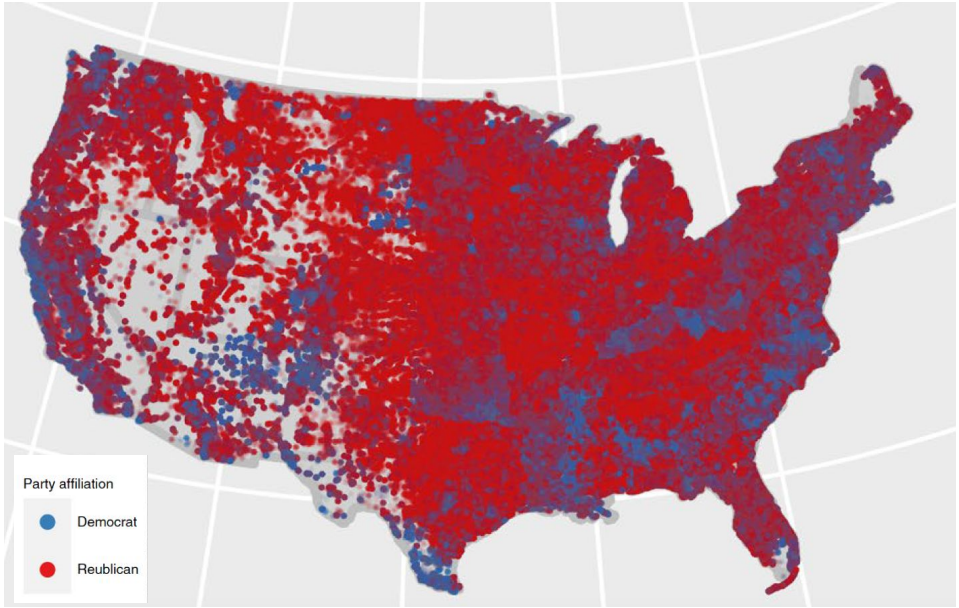
Center for Geographic Analysis
Harvard University

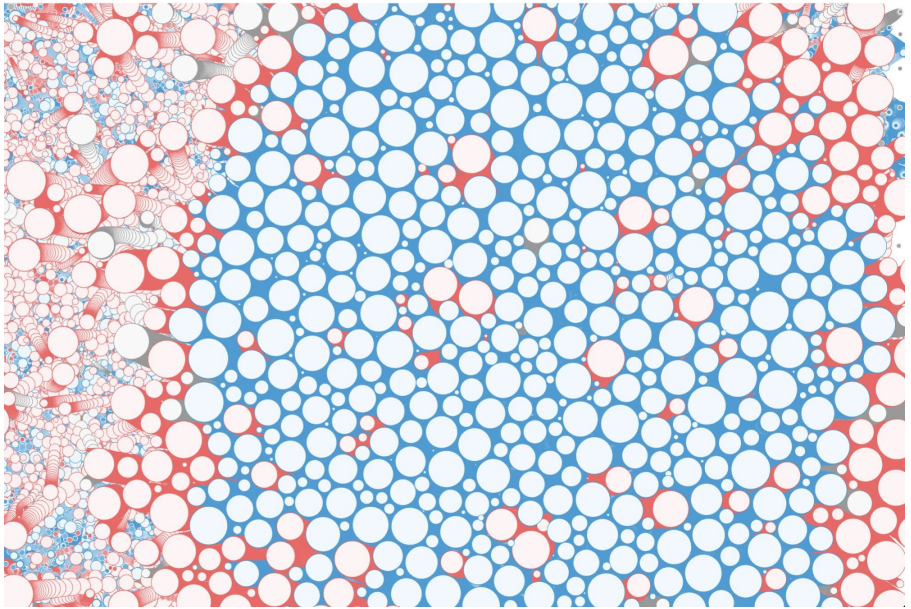
Spatiotemporal Thinking, Computing and Applications
STC

IQSS
The Institute for Quantitative Social Science

Project Overview

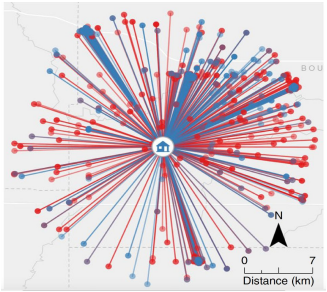
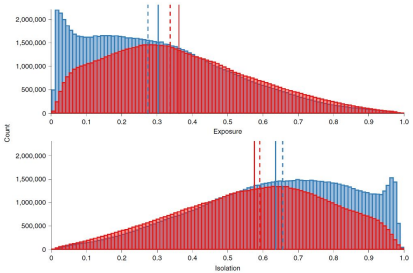
Objective: Measure partisan segregation at individual level for 180 Million U.S. voters





This work is partially sponsored by NSF Awards #1841403 and OmniSci

Challenges: Big geospatial data processing

	Traditional method	Challenges
K-Nearest Neighbor search 	• Creating pairwise distance matrix and sorting on distance • Buffer method	• Dataset of 180 Million records implies **trillion calculations** • Traditional method are slow and inefficient
Partisan exposure calculations 	• Execute using scripting language e.g., Python, R	• Partisan weight calculations from **180 Billion** distances • Scripting methods are slow and resource intense • I/O speed is very slow

[Images: Nature Human Behaviour]

Solution: Available Computing Resources

FAS RESEARCH COMPUTING
HARVARD UNIVERSITY
FACULTY OF ARTS & SCIENCES

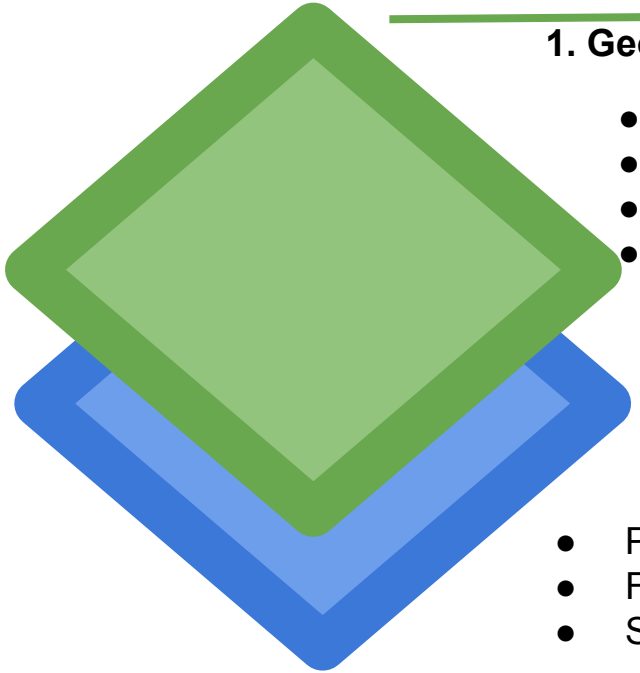
PostGIS
Spatial PostgreSQL

+

omnisci

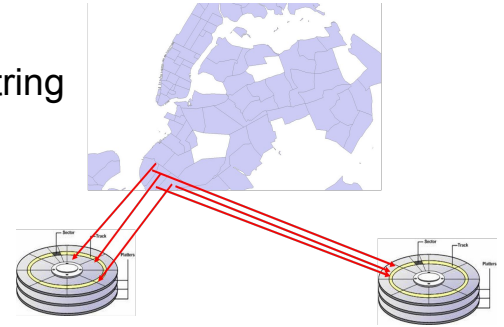
I/UCRC for Spatiotemporal Thinking, Computing, and Applications

Solution: Two-Layered Approach for KNN Calculations



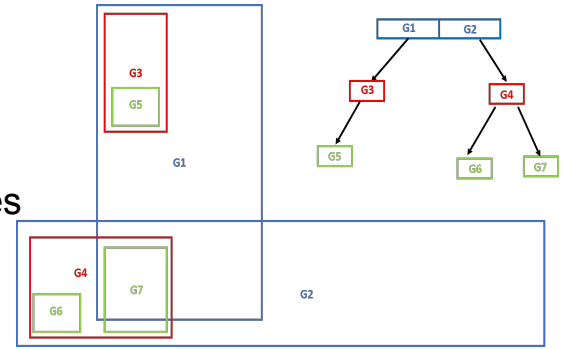
1. Geohash based Spatial Clustering

- Coordinates expressed as alphanumeric string
- Longer the shared prefix, closer the points
- Closer points clustered together on disk
- Provides fast and efficient access to data

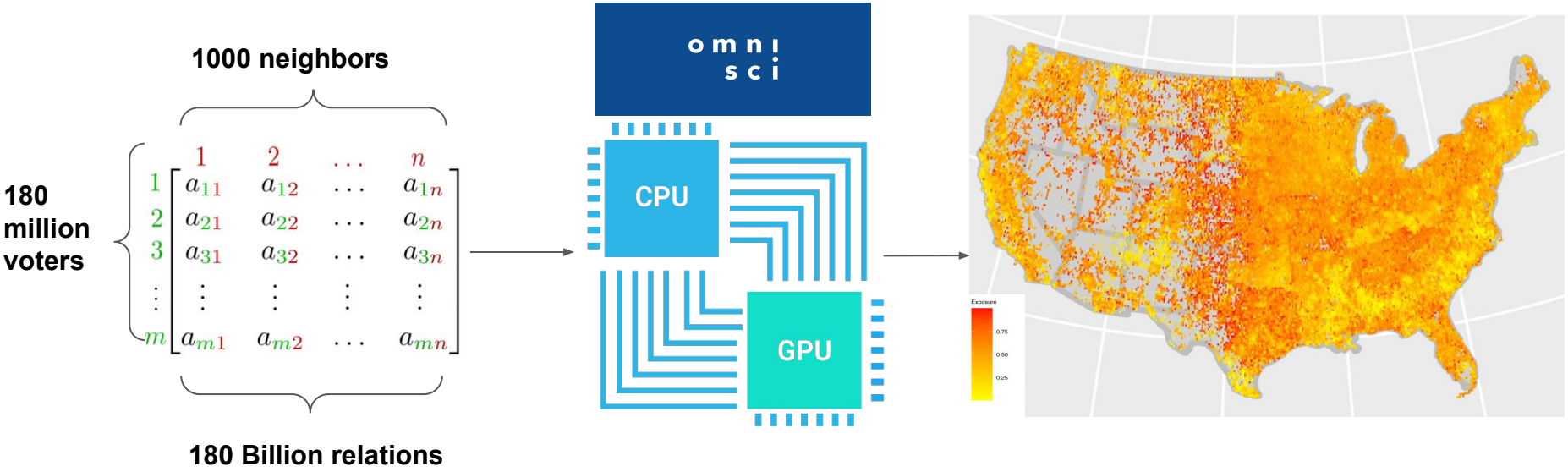


2. R-tree based Index Search

- Pure spatial Index based search
- Faster, cheaper and more efficient
- Searches up and down the Bounding boxes



Solution: Accelerated GPU based processing of partisan exposure



Solution: Novelty of our approach

Big geospatial data using Data Science

KNN calculations **@ 200,000 distances/sec**

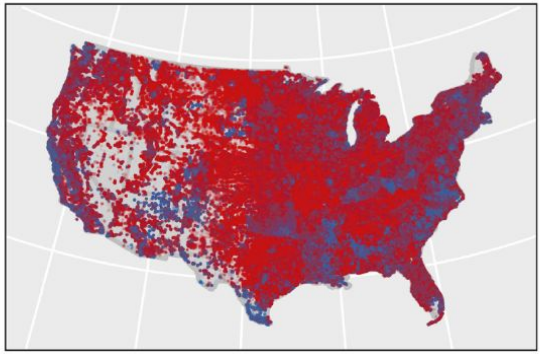
Partisan weights **@ 800M distances/sec**

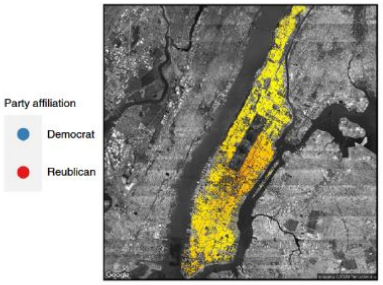
Cost and Time Effective

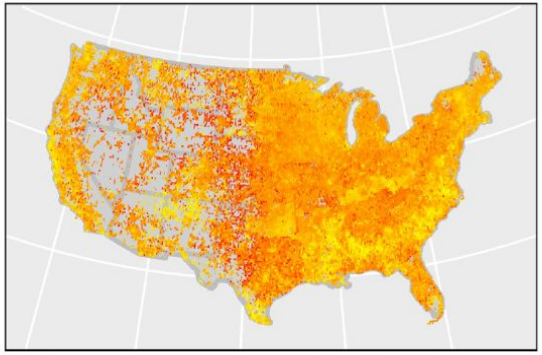
Extremely fast I/O on big data

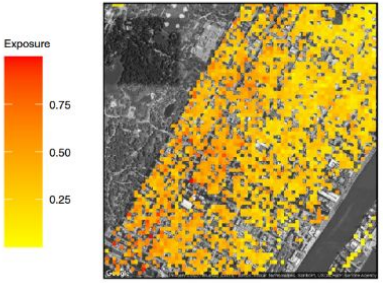
Accelerated GPU processing

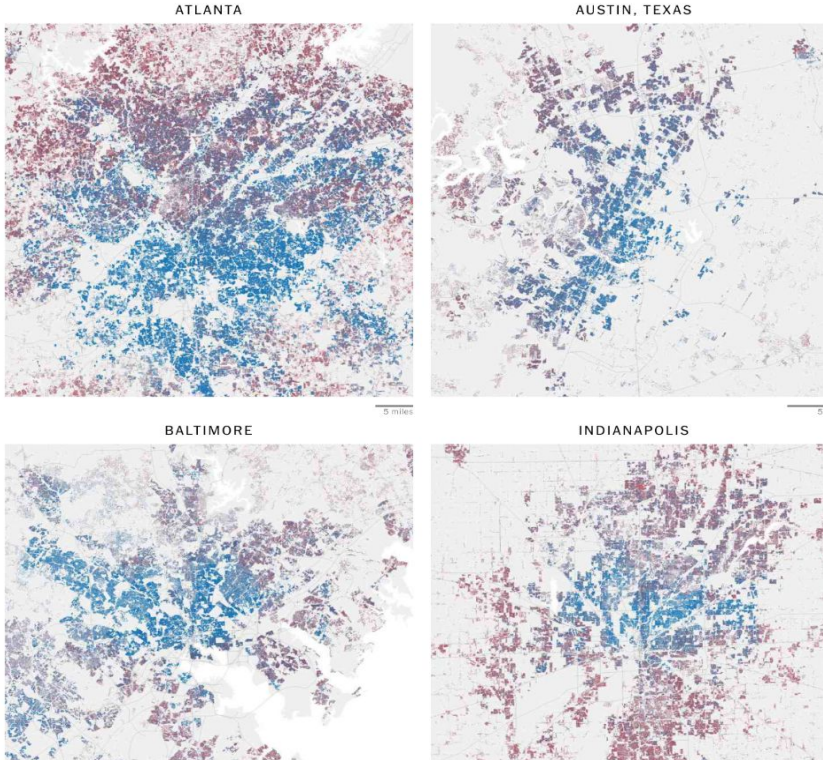
Results: Partisan exposure of individual US voters











[Images: Nature Human Behaviour]

Results: Publications and news coverage

nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01066-z

Check for updates

The measurement of partisan sorting for 180 million voters

Jacob R. Brown [1,2] and Ryan D. Enos [1,2]

Segregation
maladies. In
cratic goverr
the local resi
than 180 mil
proportion o
levels of par
tion. Moreov

Segregati
a range
group co
democratic g
Segregation i
the social scie
relations[5,6], th
and political
using aggrega
the United St
describe stark
political parti
and threateni
association be
claims of incr

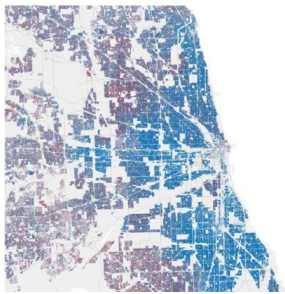
The New York Times

TheUpshot

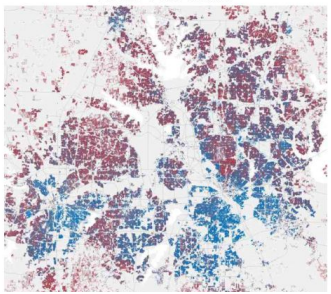
A Close-Up Picture of Partisan Segregation, Among 180 Million Voters

By Emily Badger, Kevin Quealy and Josh Katz March 17, 2021

CHICAGO



DALLAS-FORT WORTH



FAST COMPANY

03-13-21

Your partisan filter bubble is now following you around in the real world

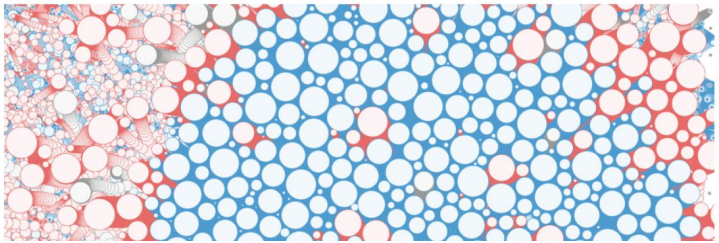
Not only d
find they d

The New York Times

Do You Think You Live in a Political Bubble?

Republicans and Democrats are increasingly cut off from one another, rhetorically and geographically. Do we need to pop our bubbles?

47



References

[1] Brown J. & Enos R., The measurement of partisan sorting for 180 million voters, Nature Human Behavior, 2021 https://www.nature.com/articles/s41562-021-01066-z.epdf

[2] Badger B., Quealy K. & Katz. J, A Close-Up Picture of Partisan Segregation, Among 180 Million Voters, The New York Times, 2021 https://www.nytimes.com/interactive/2021/03/17/upshot/partisan-segregation-maps.html

[3] Kakkar D., Lewis B., Singh R., OmniSci Virtual Summit, 2020 https://www.youtube.com/watch?v=3DlOeWqDMSs

[4] Kakkar D., Lewis B., Scaling geospatial processes on Harvard's high-performance cluster, Harvard DataFest, 2020 https://drive.google.com/file/d/1FEnh-okCNLuthtyQtoBldyid7D6Sb-F_/view?usp=sharing

[5] Introduction to Cluster Computing on FASRC: https://www.rc.fas.harvard.edu/wp-content/uploads/2019/12/Intro-to-Cannon.pdf

[6] About Postgis: https://postgis.net/

[7] OmniSci Overview: https://docs.omnisci.com/latest/4_distributed.html