The background image shows an aerial view of the University of Wisconsin-Madison campus during sunset. The sky is filled with warm orange and yellow hues. In the foreground, the calm waters of Lake Mendota are dotted with many small sailboats. The university buildings, featuring various architectural styles and colors like red brick and white stone, are visible along the shoreline and across the water.

Exploring the World of Geospatial Data Science

Dr. Song Gao
Geospatial Data Science Lab
University of Wisconsin-Madison



Email: song.gao@wisc.edu

<https://geods.geography.wisc.edu/>

An aerial photograph of a city skyline at sunset, with a large body of water in the foreground. The sky is filled with warm orange and yellow hues from the setting sun. The city buildings are visible along the shore, and many small sailboats are scattered across the water.

Slides, Data and Code:
https://github.com/GeoDS/geods_workshop





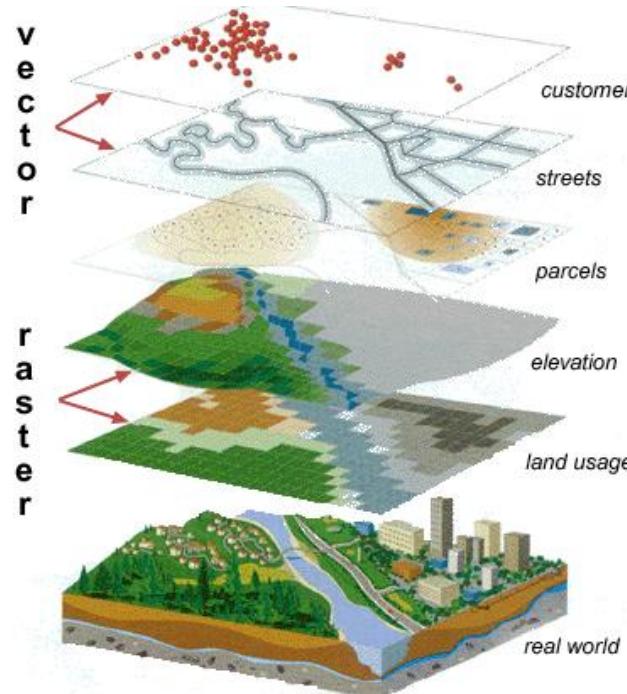
(Geo-)Spatial Data Science

- Spatial data science can be viewed as a subset of “data science” that focuses on the characteristics of spatial data, i.e., the importance of “**where**.”
- Spatial data science treats **location**, **distance**, and **spatial interaction** as core aspects of the data.
- Spatial data science employs specialized methods and software to **store**, **retrieve**, **analyze**, **visualize** and **learn** from such data.



Geospatial Data Model

- A logical data model is how data are organized for use by the GIS.
- GIS have traditionally used either:
 - Vector (points, lines, polygons)
 - Raster (pixels)

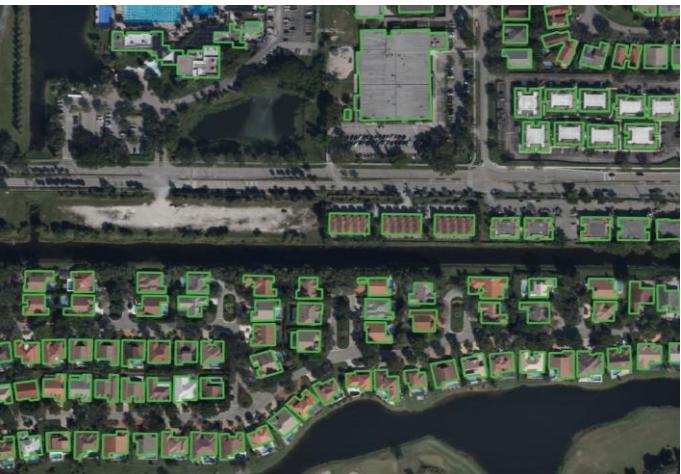
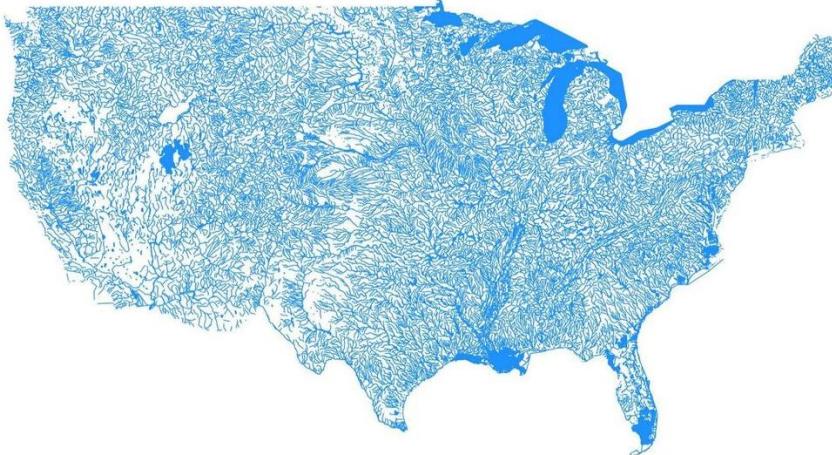
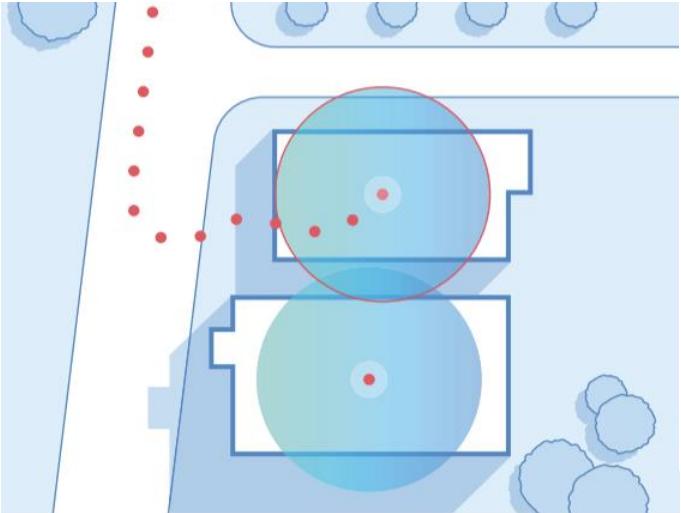


The raster view of the world	Happy Valley spatial entities	The vector view of the world
	Points: hotels	
	Lines: ski lifts	
	Areas: forest	
	Network: roads	
	Surface: elevation	

Types of Spatial Data



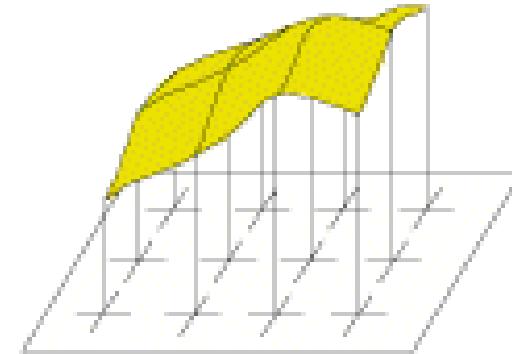
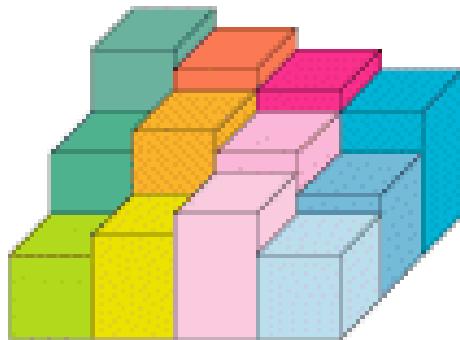
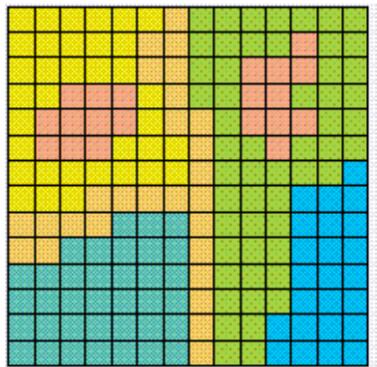
- Points, Lines, Polygons, and Networks



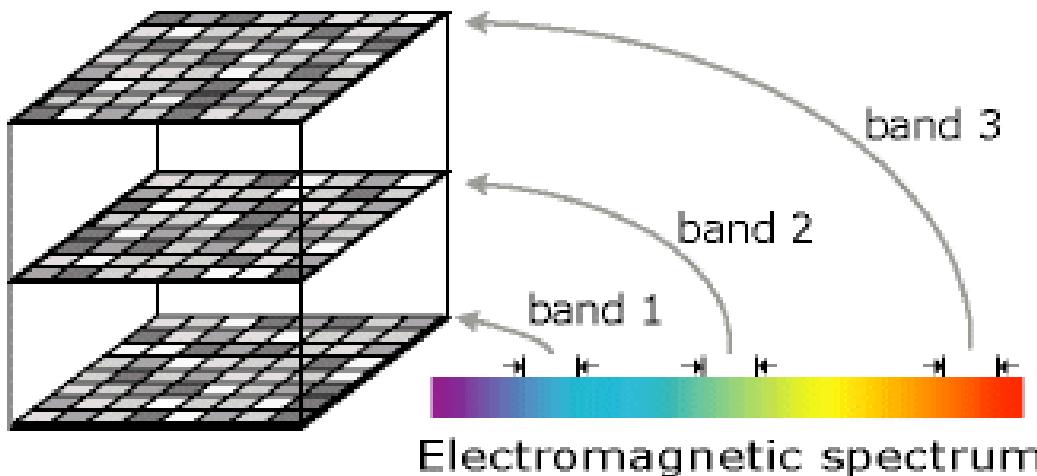
Types of Spatial Data



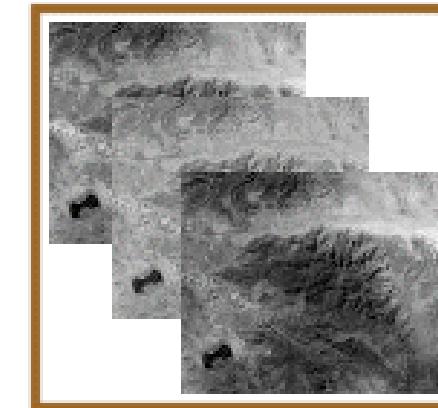
- Raster



+	315	+	319	+	321	+	323
+	317	+	323	+	320	+	326
+	313	+	318	+	325	+	323



Multiband
raster dataset



RGB composite



Google Earth Engine Data Catalog


**Landsat and
Sentinel**

Raw, TOA, SR, ...


MODIS

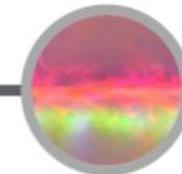
Daily, NBAR, LST, ... SRTM, GTOPO, NED, ...


Terrain

SRTM, GTOPO, NED, ...


Land Cover

GlobCover, NLCD, ...


Atmospheric

NOAA NCEP, OMI, ...

... and many more, updating daily!

> 200 public datasets

> 5 million images

> 4000 new images every day

> 5 petabytes of data

Google Earth Engine

Google Earth Timelapse

About the project →

Earth Timelapse is a global, zoomable video that lets you see how the Earth has changed over the past 37 years.

Explore Timelapse in 3D using [Google Earth](#).

Timelapses around the world



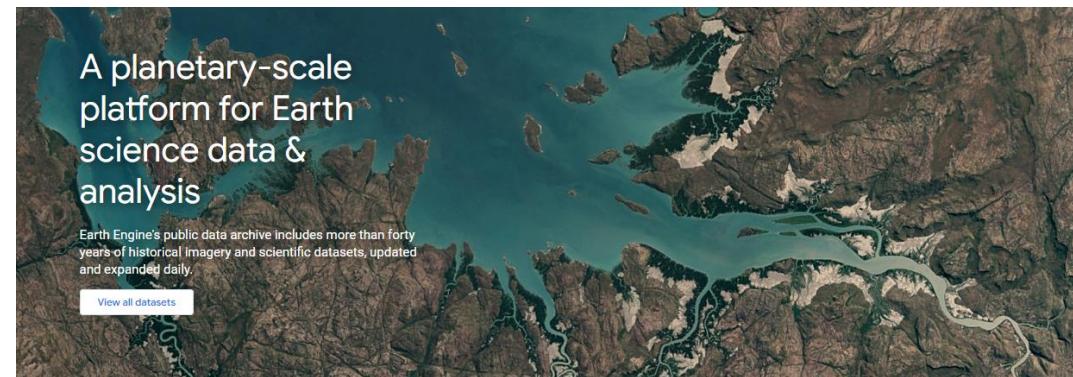
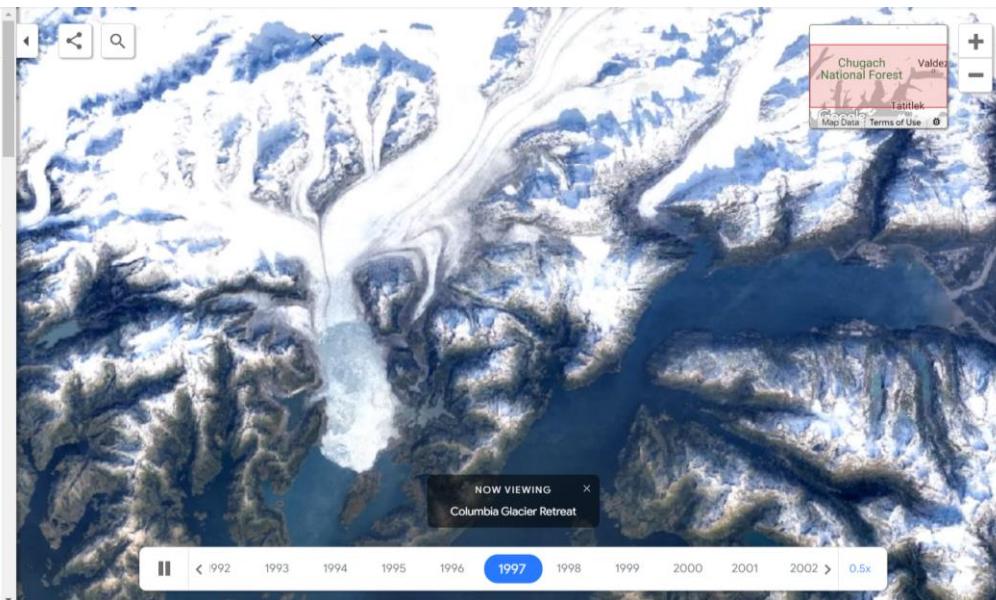
Columbia Glacier Ret...
Alaska, USA



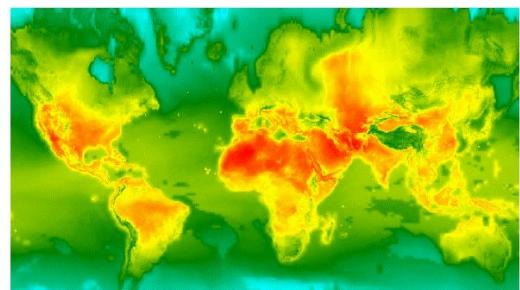
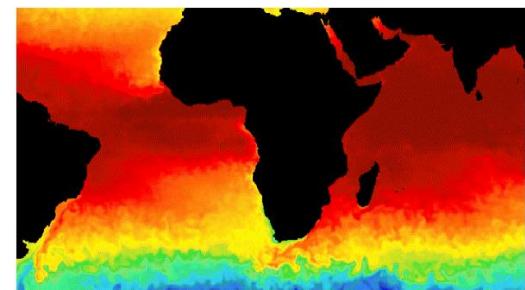
Mining
Alberta, Canada



Construction of the ...
Schonefeld, Germany



Climate and Weather



<https://developers.google.com/earth-engine/datasets>

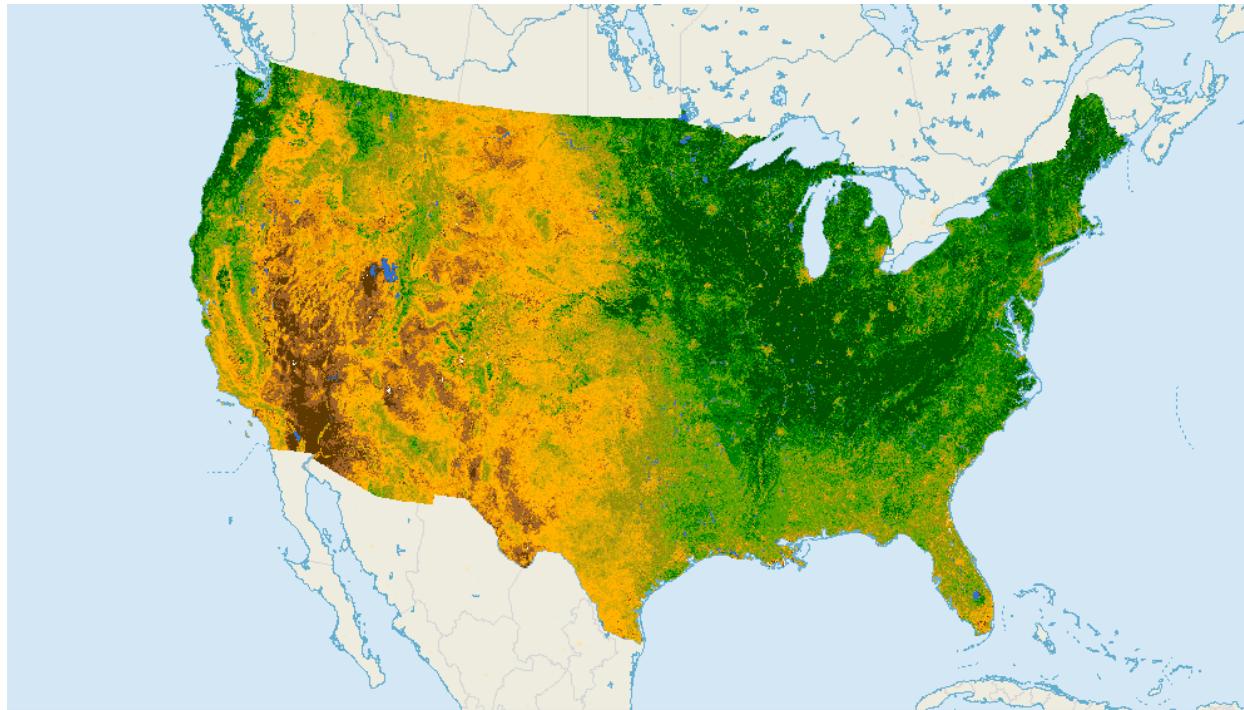
<https://earthengine.google.com/timelapse/>

Geocomputation



Normalized Difference Vegetation Index

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$



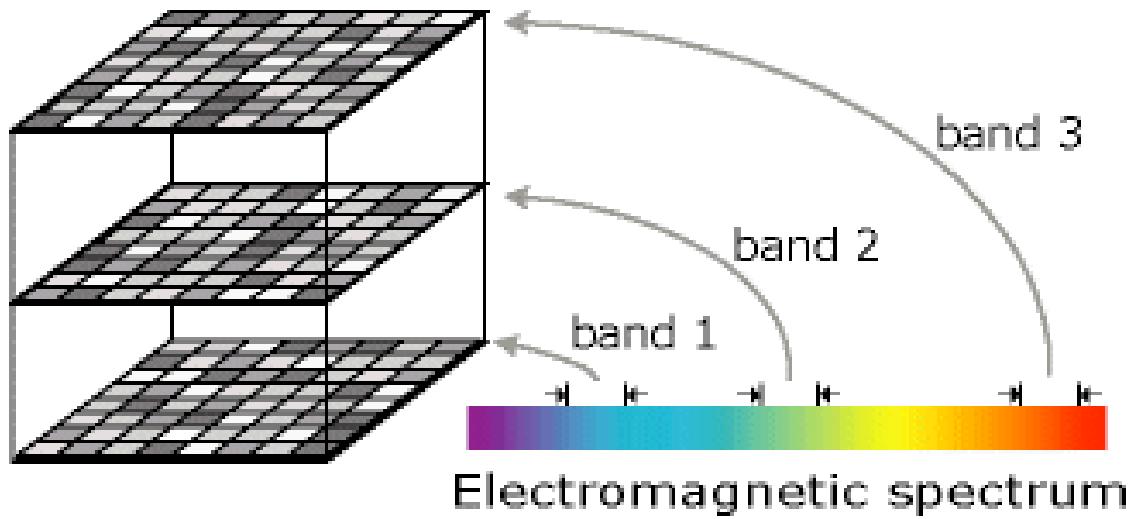
Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

https://developers.google.com/earth-engine/tutorials/tutorial_api_06

Classification



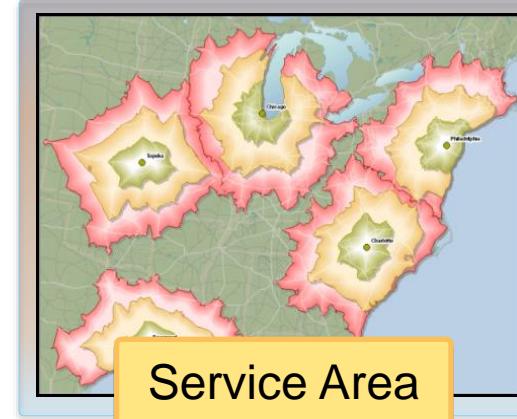
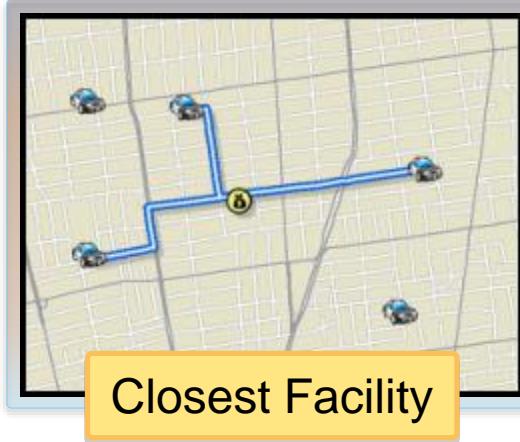
- Utilize **spatial and spectral features** in remote sensing images to assign **a class label** into each pixel.



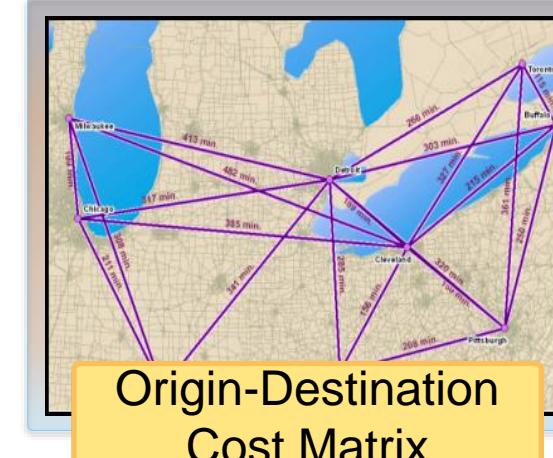
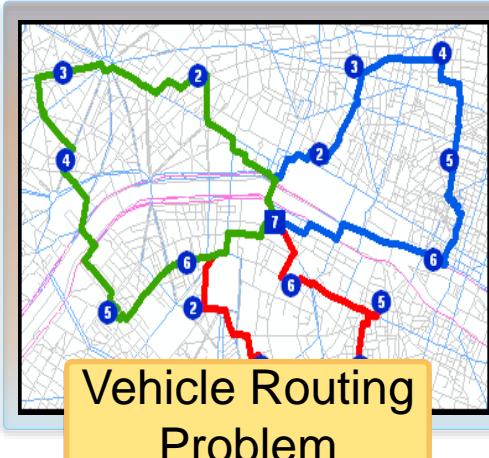
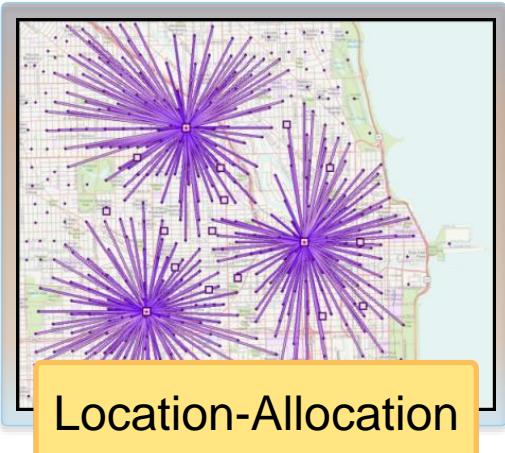
Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Street Network Analysis

GeoDS Lab @UW-Madison

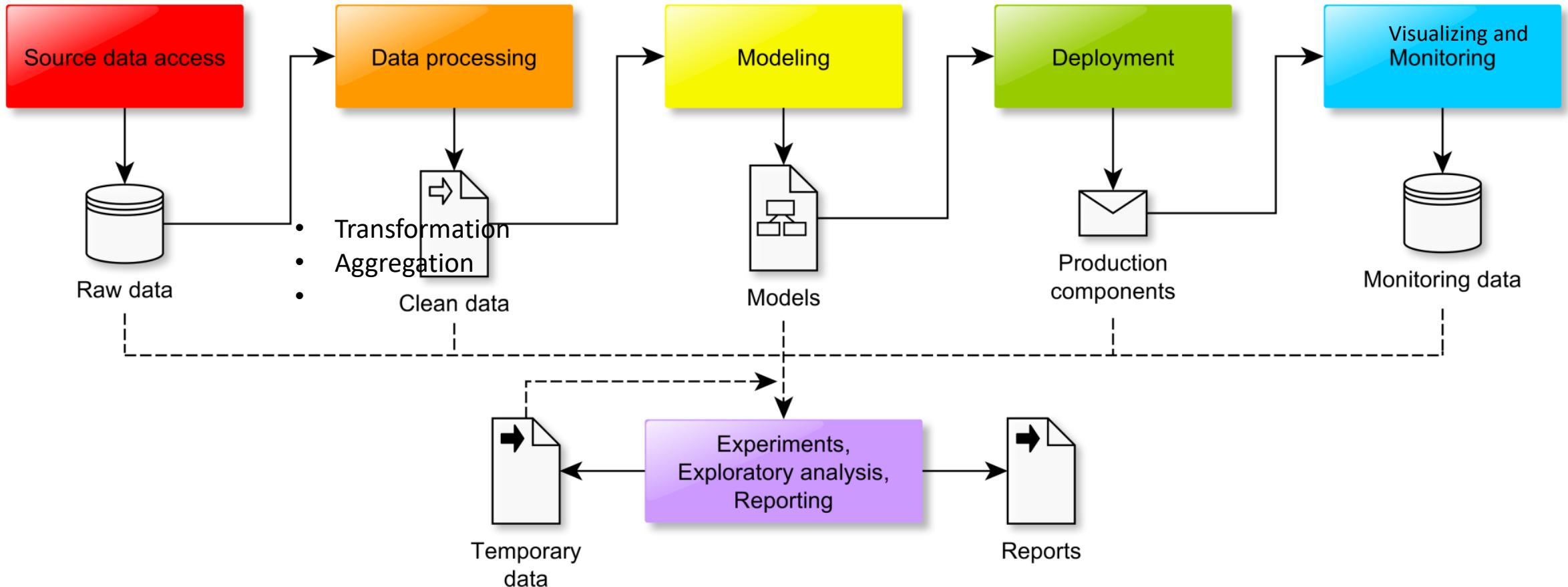


ArcGIS Network Analyst Solving transportation problems



Workflow for (Spatial) Big Data Analytics

GeoDS Lab @UW-Madison

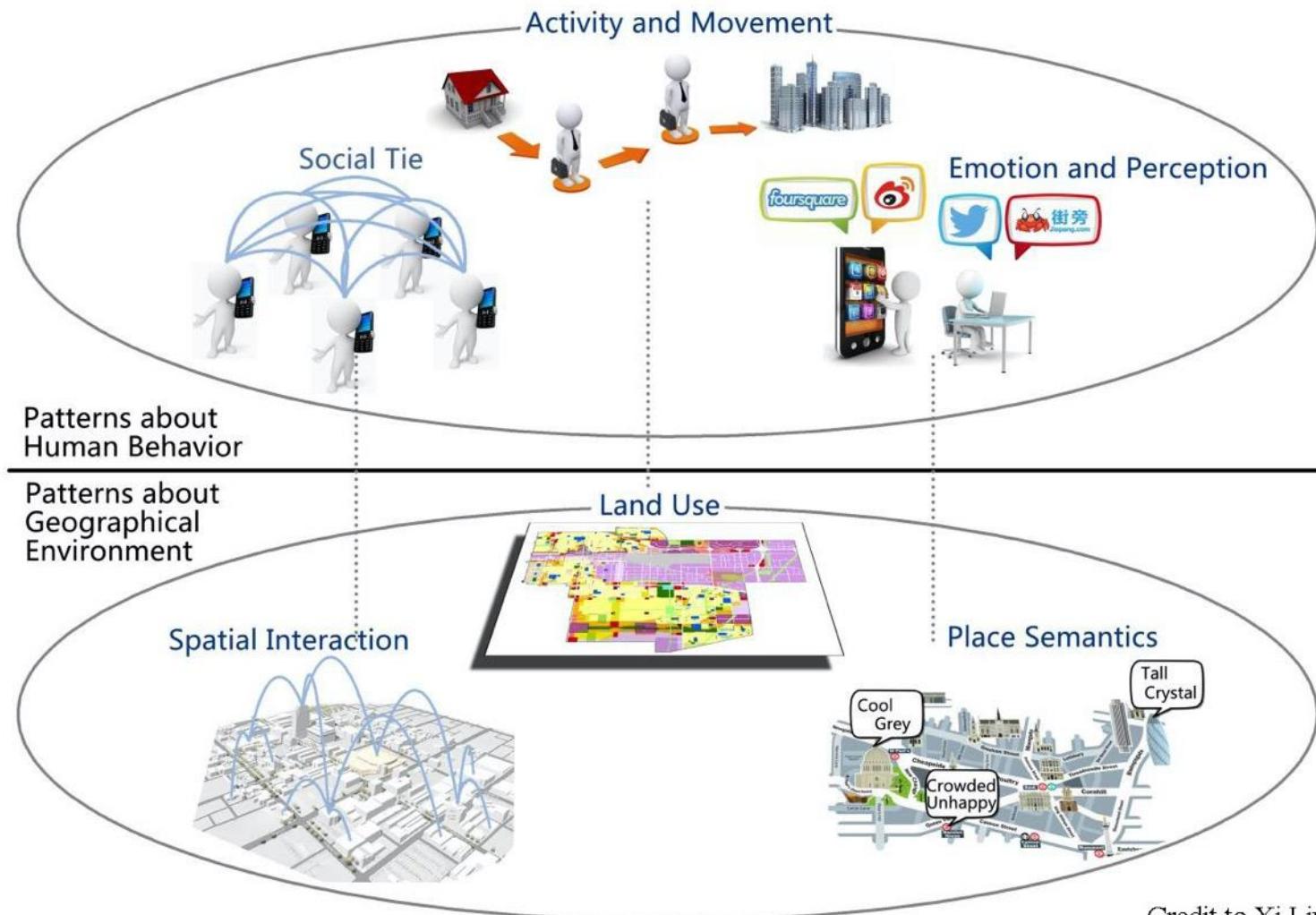


Data-Driven Human Mobility Research

GeoDS Lab @UW-Madison



- Trip Surveys
- GPS trajectories
- Call detail records (CDRs)
- Social media
- Location-based apps
- Telematics data



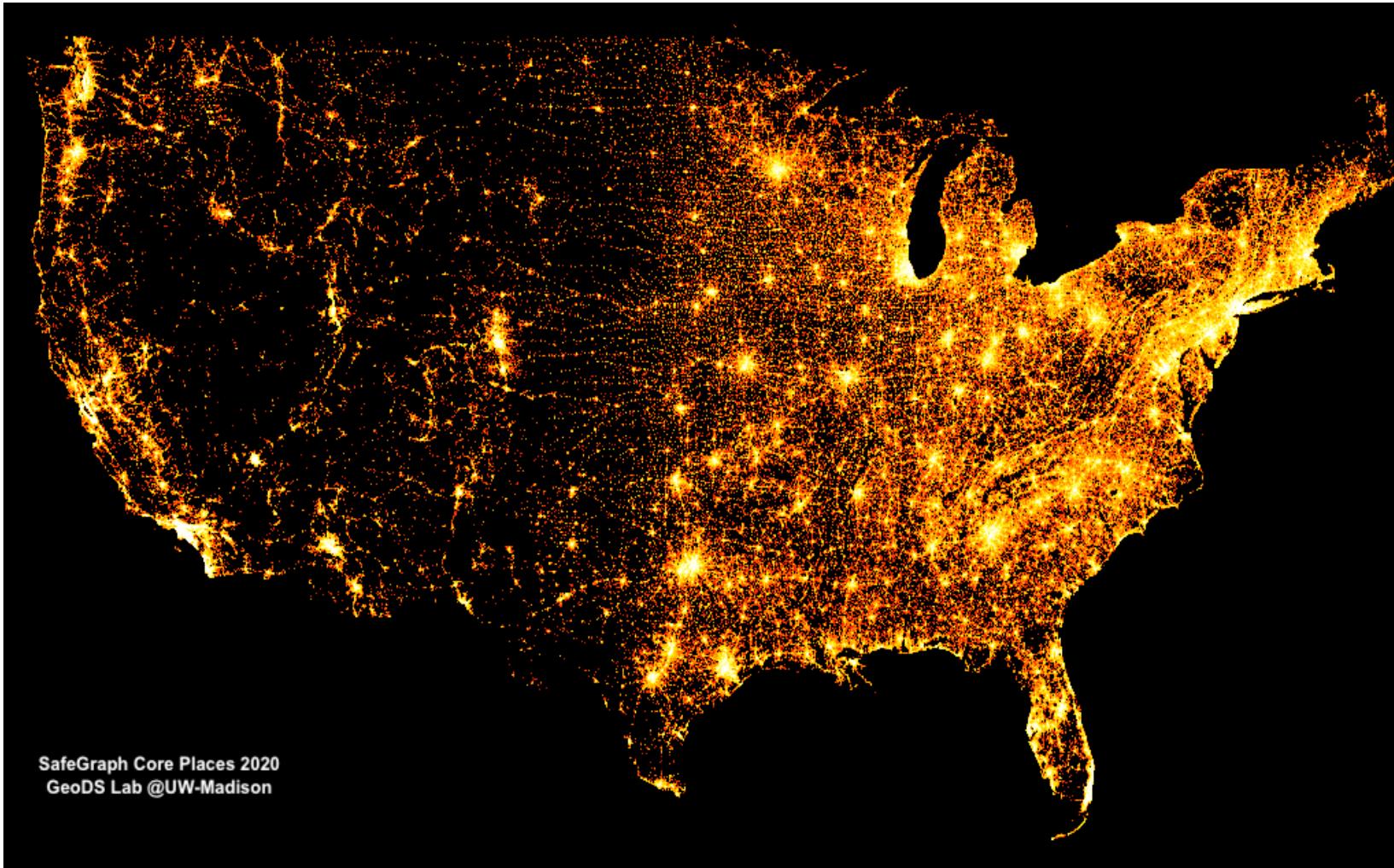
Credit to Xi Liu

Human Mobility Tracking from Anonymous Mobile Phone Devices

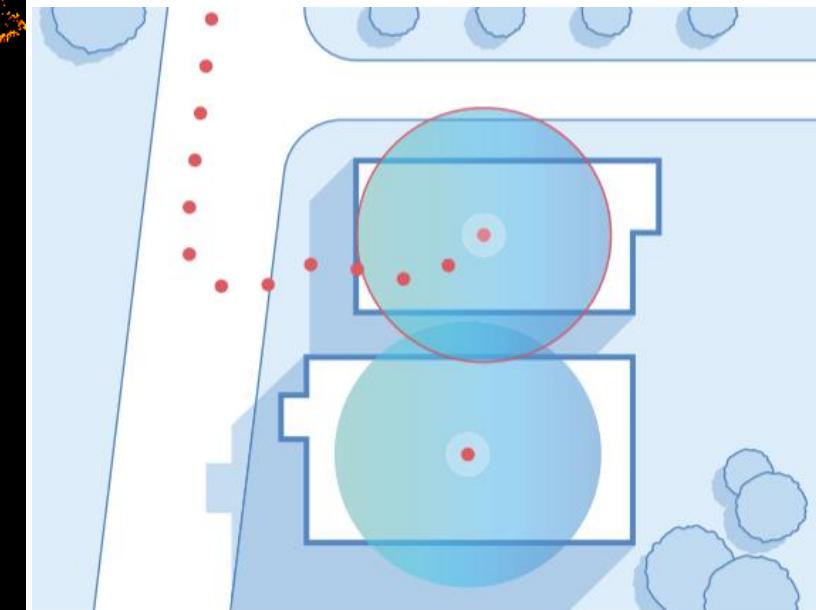
GeoDS Lab @UW-Madison



- 45M+ devices; 10M+ POI; 8,500+ Brands; 400 Place Categories



SAFE GRAPH



Core Places



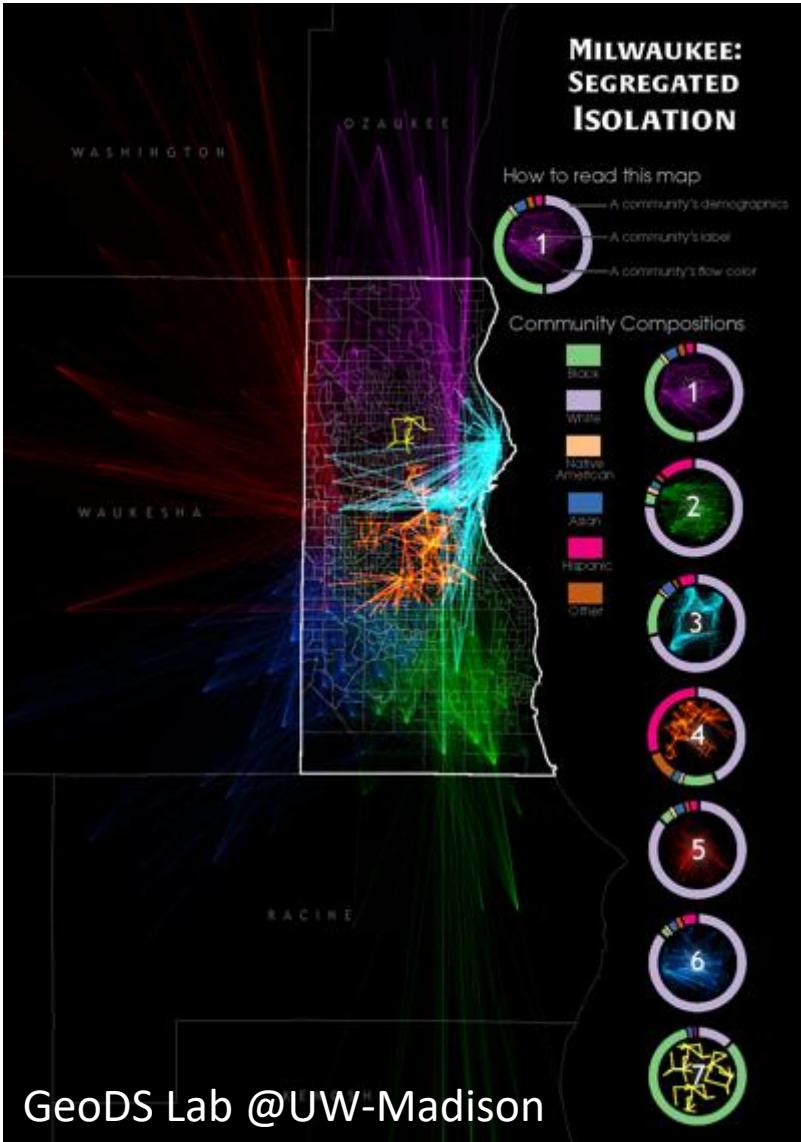
Geometry



Patterns

Identifying Socio-demographic Isolation

GeoDS Lab @UW-Madison



- Over half-million mobile devices location pings for one-month
- Analyze Human mobility spatial-social interactions
- Using community detection algorithm on spatial networks
- Identify isolated communities that have less frequent interactions outside of their residence neighborhoods.

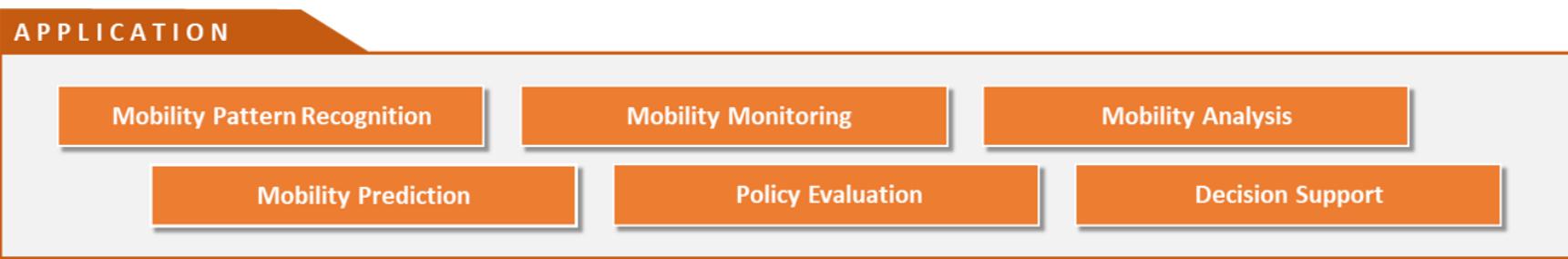
Timothy Prestby, Joseph App, Yuhao Kang, Song Gao. (2019) Understanding Neighborhood Isolation through Spatial Interaction Network Analysis using Location Big Data. *Environment and Planning A: Economy and Space*. 52, 6.

Human Mobility Tracking from Anonymous Mobile Phone Devices

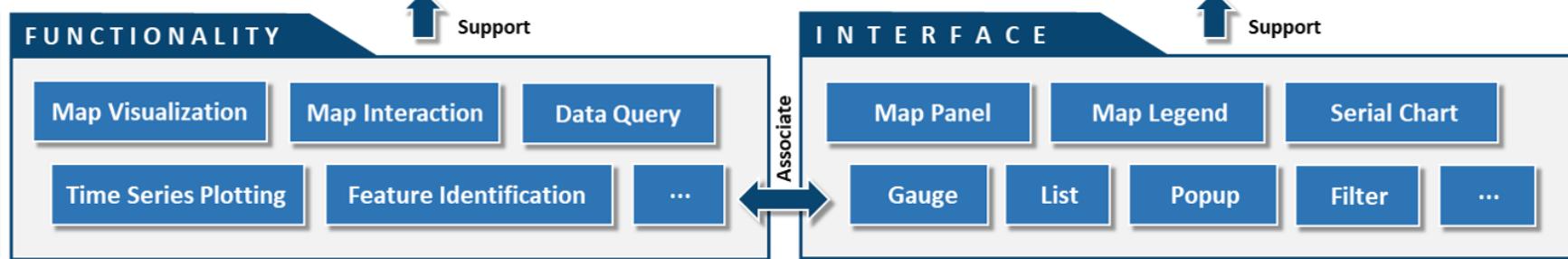
GeoDS Lab @UW-Madison



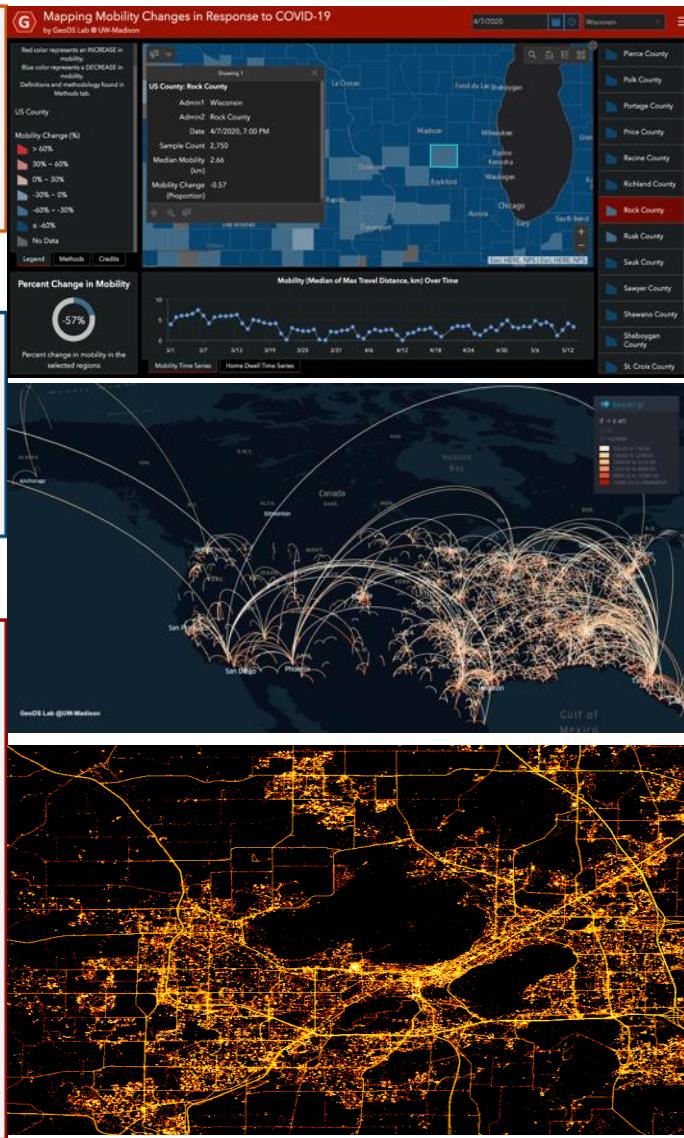
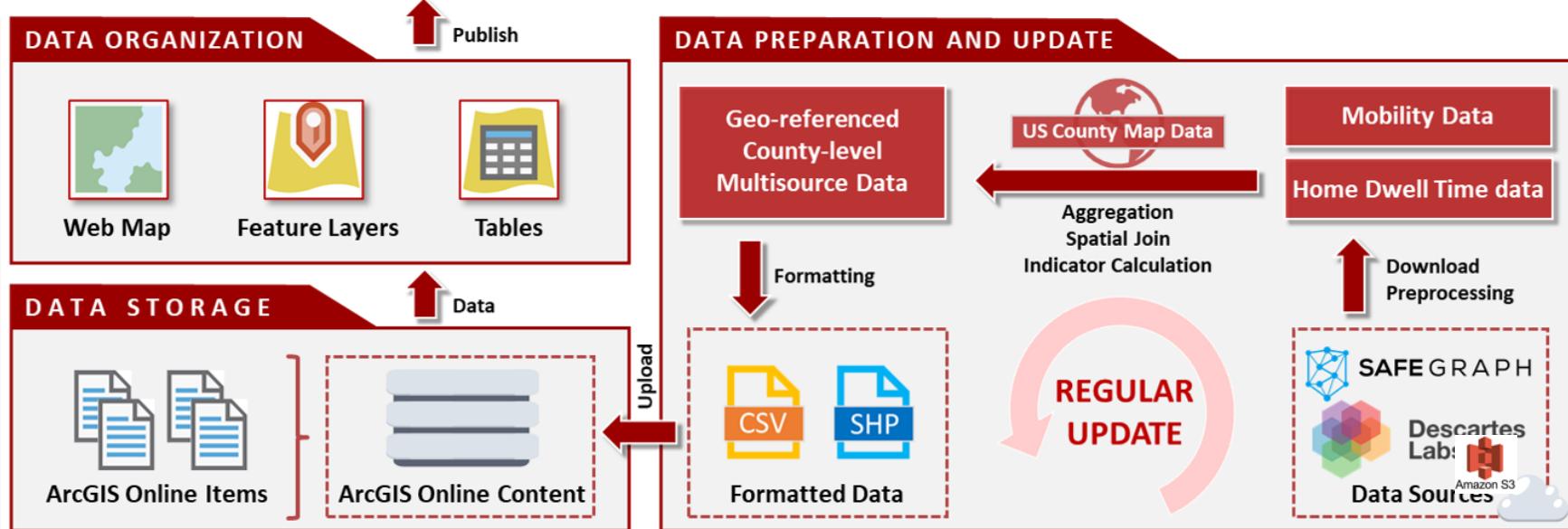
Application



Methodology

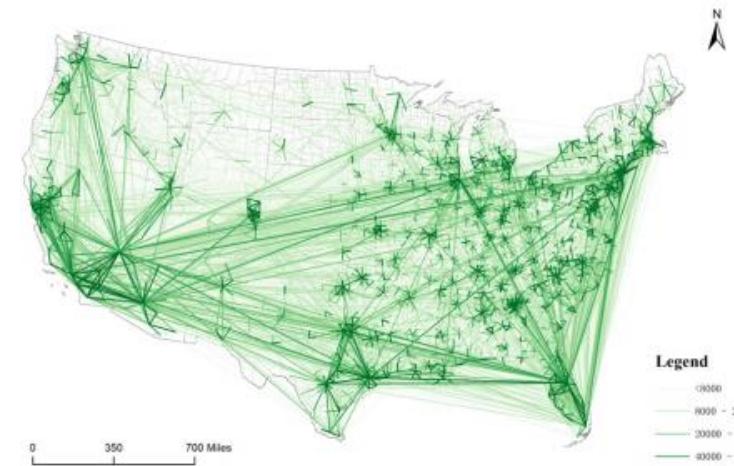


Data and Features

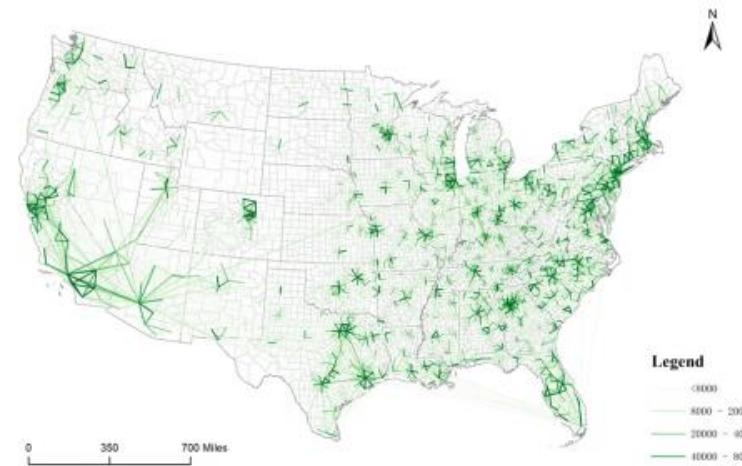


Multiscale Dynamic O-D Flow Open Data

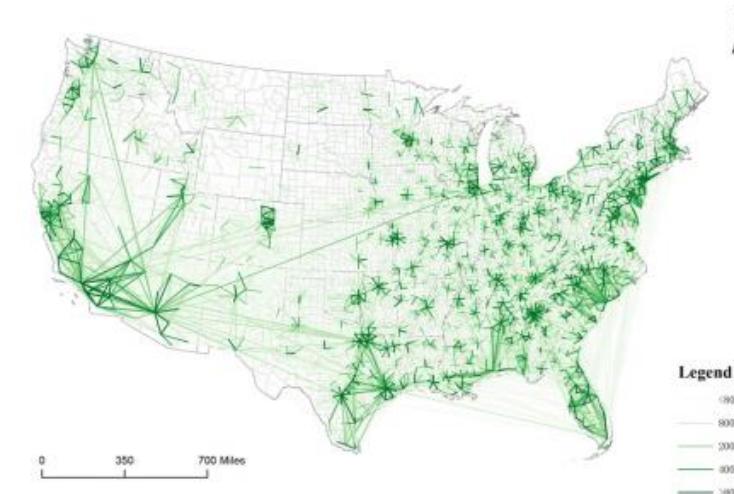
Weekly Population Flows between March 2nd and March 8th



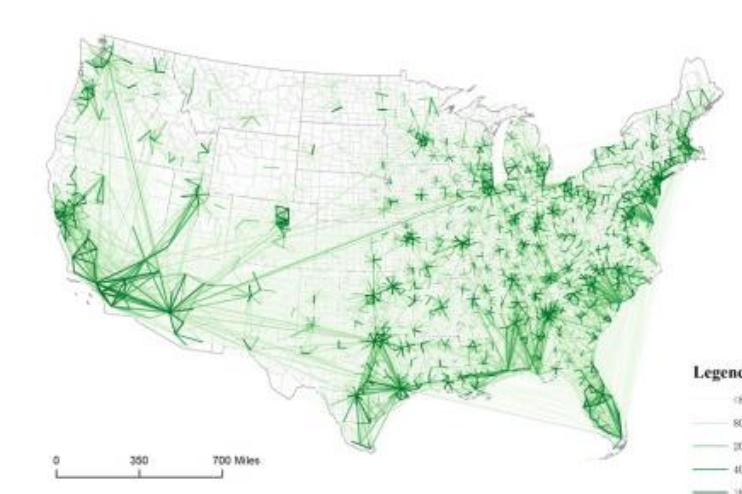
Weekly Population Flows between April 6th and 12th



Weekly Population Flows between May 11th and May 17th



Weekly Population Flows between May 25th and May 31st



Geospatial Data Science Lab
UW-Madison



SAFE GRAPH

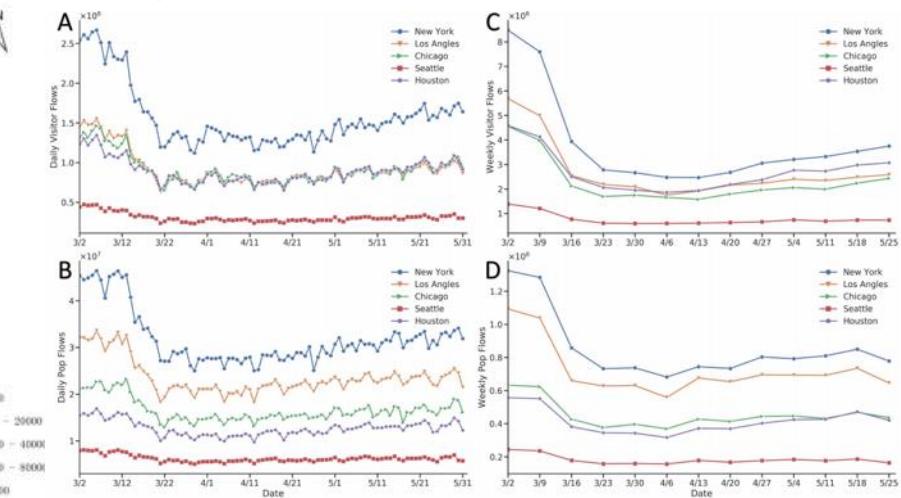


Multiscale Dynamic Human Mobility Flow Dataset in the U.S. during the COVID-19 Epidemic

GeoDS Lab, Department of Geography, University of Wisconsin-Madison.
[Website](#) · [View Demo](#)

Table of Contents

- Citation
- About the Project
- Data Processing and Data Descriptor
- Field Descriptions

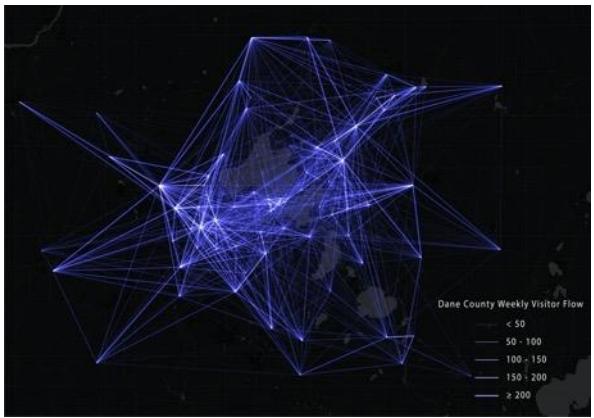


Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., & Kruse, J. (2020). Multiscale Dynamic Human Mobility Flow Dataset in the US during the COVID-19 Epidemic. *Scientific Data*. 7, 390.

<https://github.com/GeoDS/COVID19USFlows>

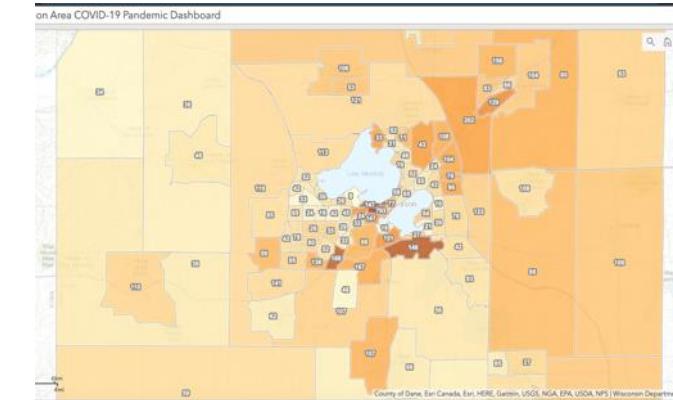
Intra-County Modeling of COVID-19 Spread

GeoDS Lab @UW-Madison

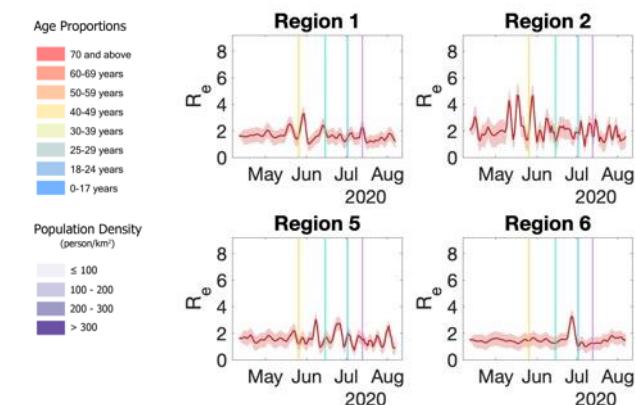
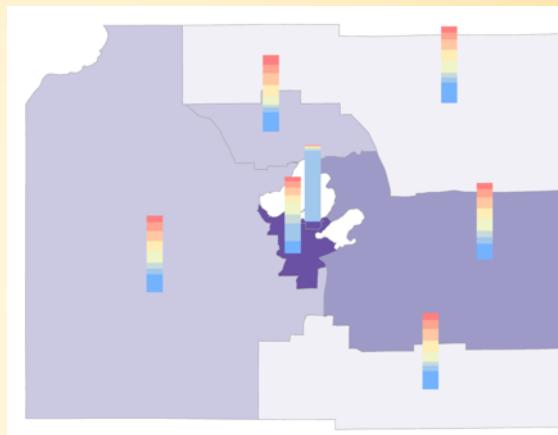


Mobility-based Modeling

Spatial Clustering SEIR model



Divide a county into multiple smaller regions according to traffic flows.



Hou, X., Gao, S.*, Li, Q., Kang, Y., Chen, N., Chen, K., Rao, J., Ellenberg, J. & Patz, J. A. (2021). Intra-county modeling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age and race. *Proceedings of the National Academy of Sciences*, 2021, 118 (24) e2020524118



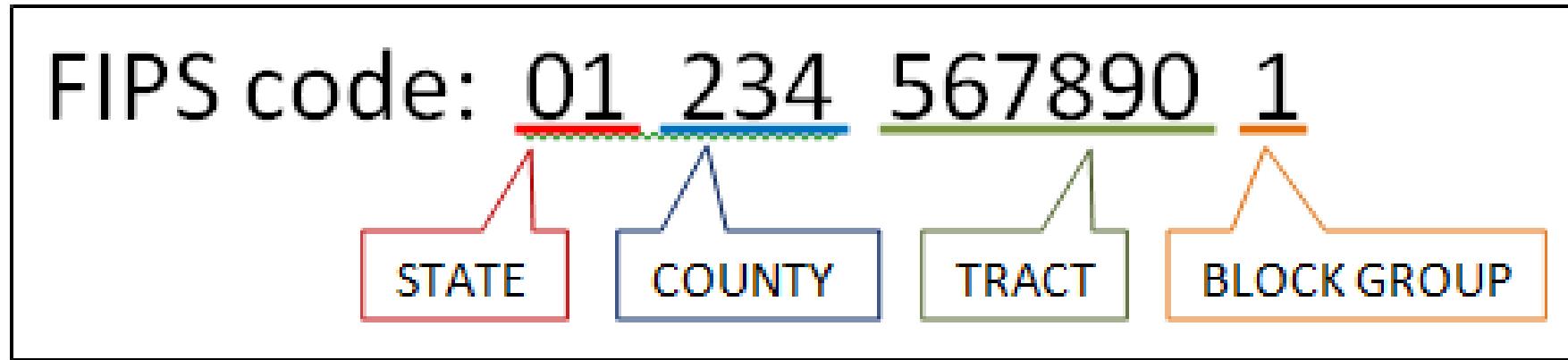
- Uber open-source geospatial analysis tool to render large datasets
 - **Layers: (2D& 3D)**
 - points
 - arcs
 - paths
 - polygons
 - grids
 - hexbins
 - **GPU- accelerated**
 - large data set



<https://kepler.gl/>



Human Mobility Analytics: Data transformation across geographical scales



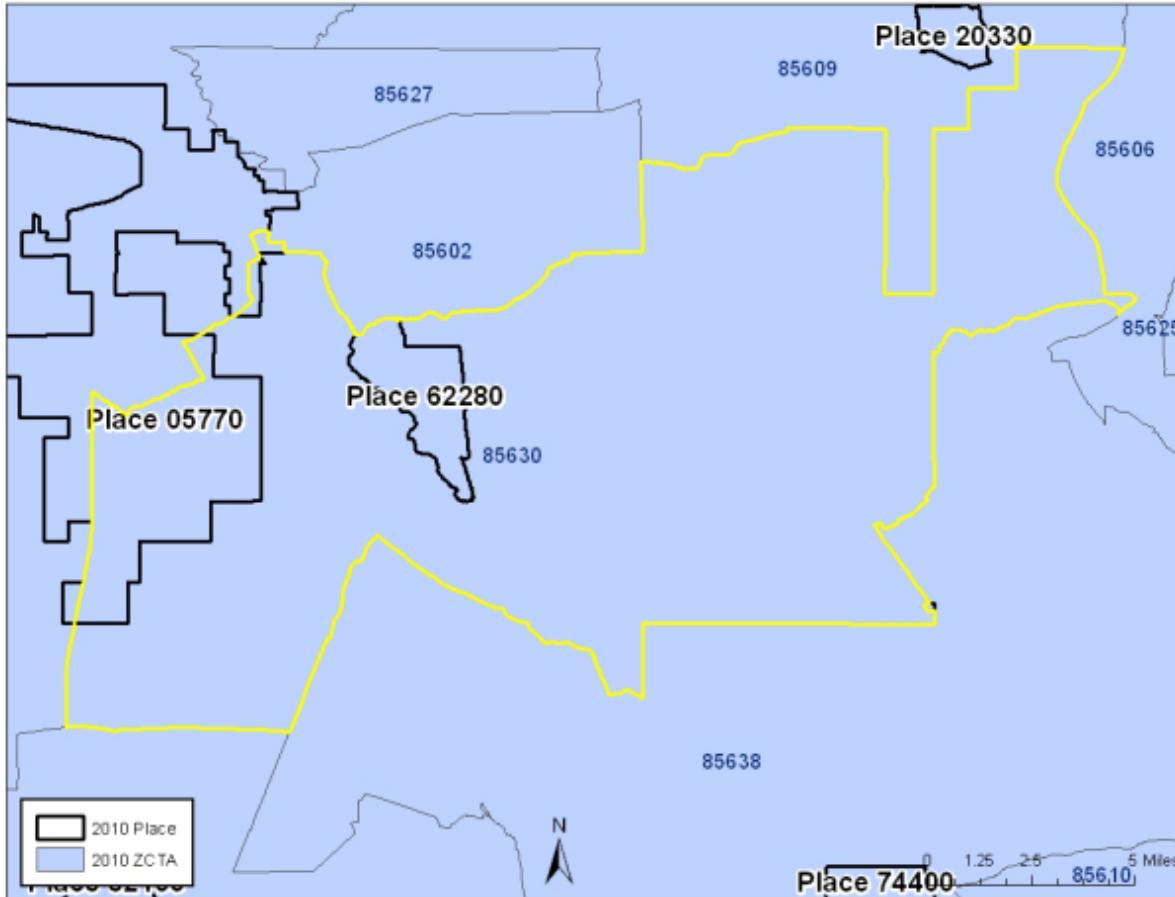
- **Federal Information Processing Series (FIPS) codes** are numbers which uniquely identify geographic areas.
- Different digits have different meanings. For example, “55” refers to the Wisconsin state, and “55025” refers to the Dane county.
- **Given that the first digit number might be 0, it is necessary to store it as a string while not an integer.** <https://www.census.gov/library/reference/code-lists/ansi.html>

Aggregate Census Tract data to Zip code

GeoDS Lab @UW-Madison



Field	Value 1	Value 2
ZCTA5	85630	85630
STATE	04	04
PLACE	62280	05770
CLASSFP	U1	C1
GEOID	0462280	0405770
POPPT	1699	3
HUPT	804	2
AREAPT	13832720	39050439
AREALANDPT	13794418	39047940
ZPOP	2819	2819
ZHU	1335	1335
ZAREA	474025278	474025278
ZAREALAND	473769825	473769825
PLPOP	1699	5105
PLHU	804	2941
PLAREA	13832720	107379833
PLAREALAND	13794418	107253646
ZPOPPCT	60.27	0.11
ZHUPCT	60.22	0.15
ZAREAPCT	2.92	8.24
ZAREALANDPCT	2.91	8.24
PLPOPPCT	100	0.06
PLHUPCT	100	0.07
PLAREAPCT	100	36.37
PLAREALANDPCT	100	36.41



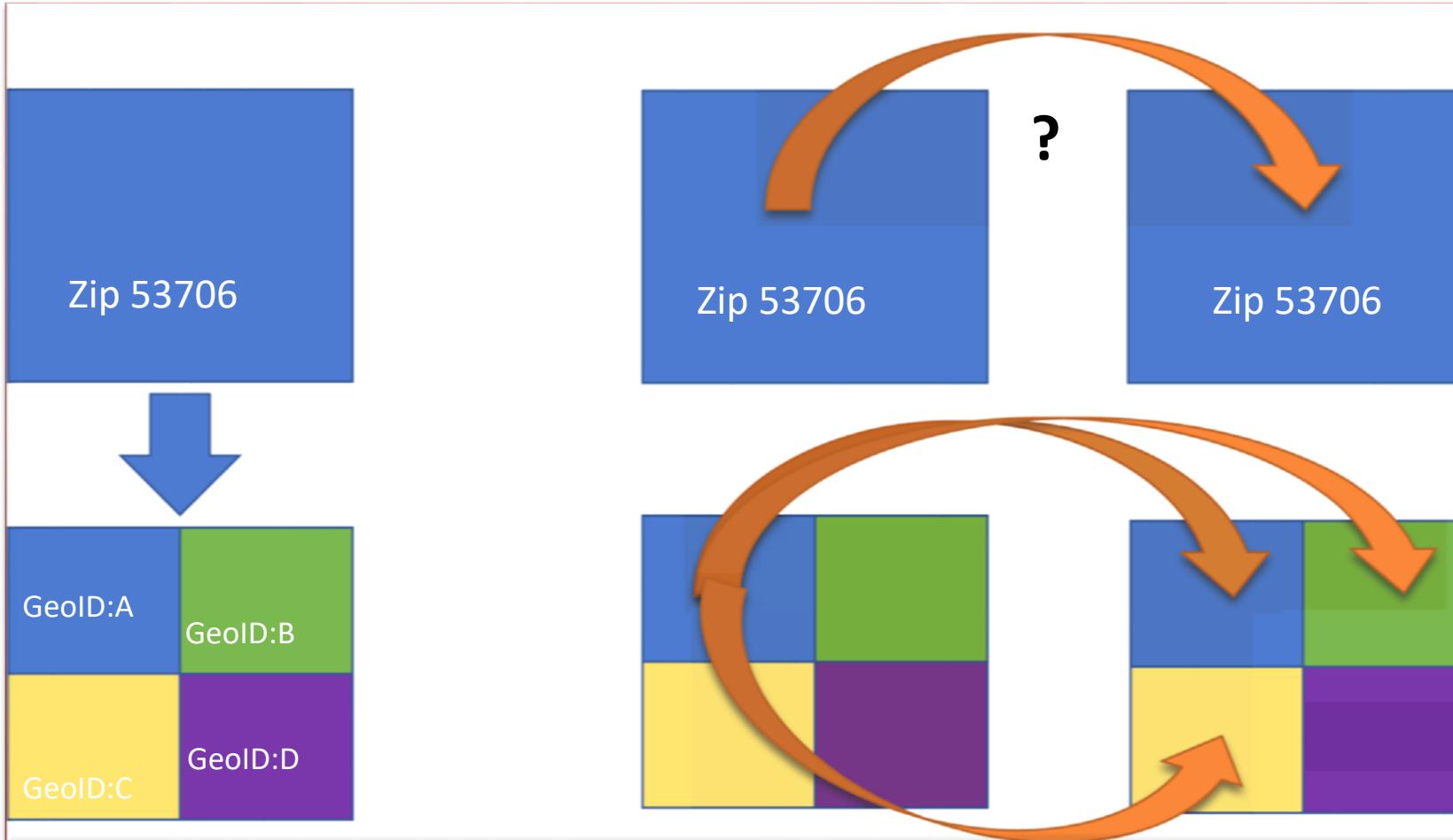
<https://www.census.gov/geographies/reference-files/2010/geo/relationship-files.html>

https://www2.census.gov/geo/docs/maps-data/data/rel/zcta_tract_rel_10.txt

Origin to Destination Flows



- From census tracts to zip code areas



Thank you!



Geospatial Data Science Lab
 UW-Madison



Song Gao
@gissong



<https://geods.geography.wisc.edu>

Twitter: @gissong

