GIS: from Data to Knowledge

for Economic Research and Business Applications

Wendy Guan & **Tao Hu**

Harvard University

Co-sponsored by

The Chinese Economic Society

China Data Institute & *Future Data Lab*

9:00-10:00 PM, August 13, 2020 (US Eastern Time)

Outline

- Why GIS
- What is GIS
- Basic properties of GIS data
- Major types of GIS Analysis
  - with examples from Harvard research projects
- Summary: the value of GIS
- Supporting COVID-19 Research with GIS Resources

The 4th Industrial Revolution

Artificial Intelligence

Big Data

Cloud Computing

Digital Disruption
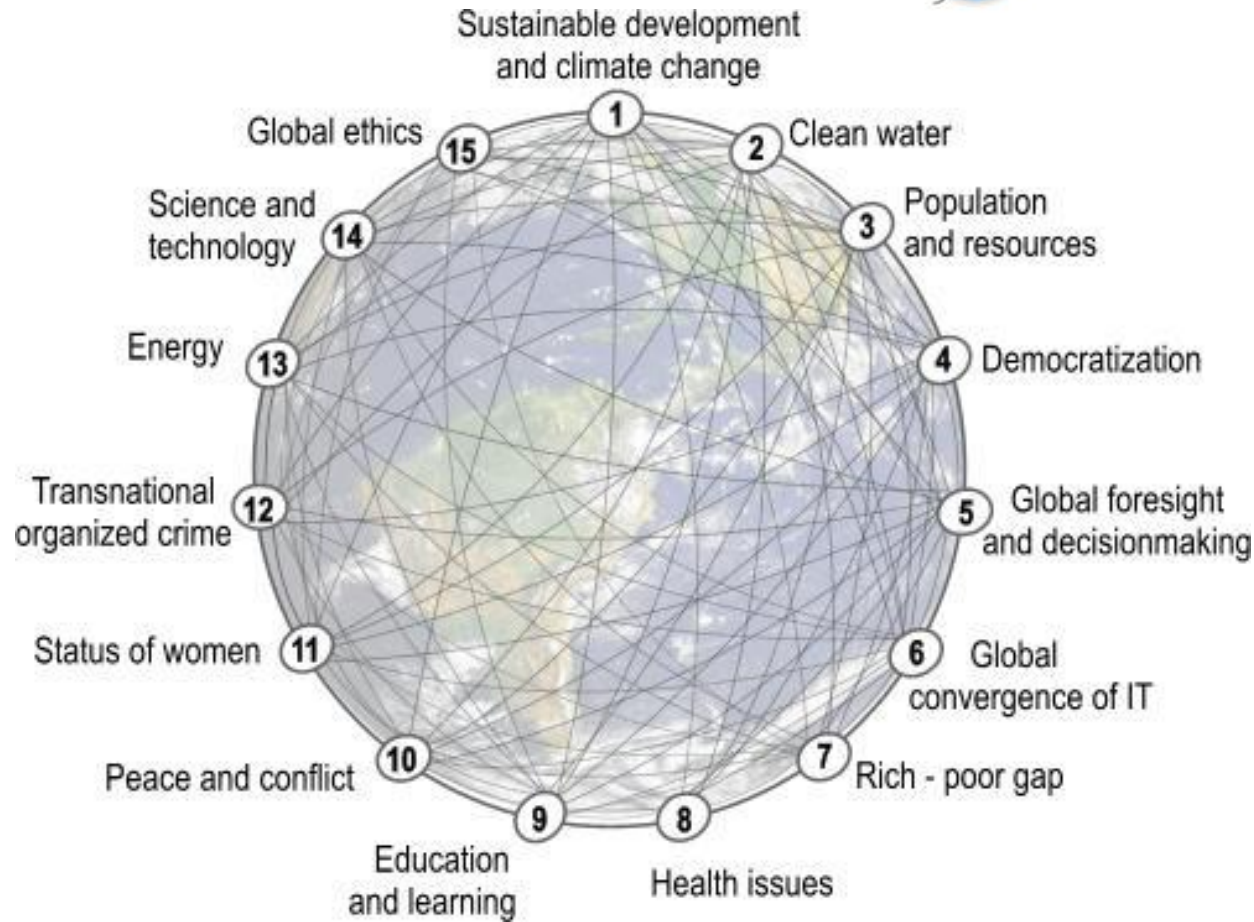
Internet of Things

Robotics

Block Chain

Geolocation becomes ubiquitous, place and time are embedded

The 15 Global Challenges

Sustainable development and climate change (1)
Clean water (2)
Population and resources (3)
Democratization (4)
Global foresight and decisionmaking (5)
Global convergence of IT (6)
Rich - poor gap (7)
Health issues (8)
Education and learning (9)
Peace and conflict (10)
Status of women (11)
Transnational organized crime (12)
Energy (13)
Science and technology (14)
Global ethics (15)

Geospatial & Temporal Data Are Everywhere

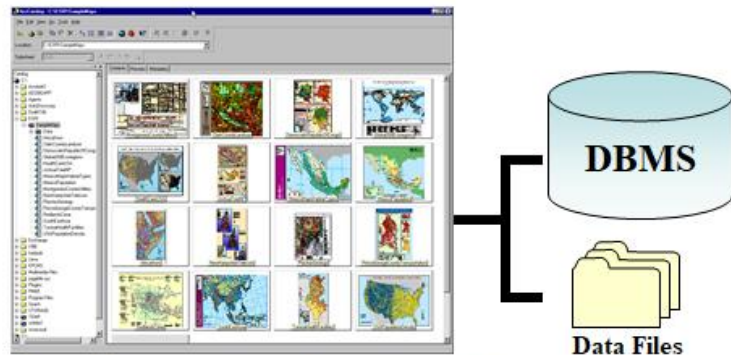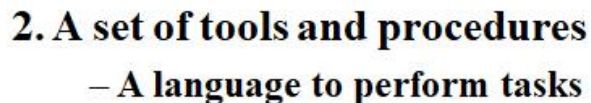Global Navigation Satellite Systems (GNSS)

Earth Observation Satellites

Smart Phones, Clothes, Cars, Homes, Cities, ...
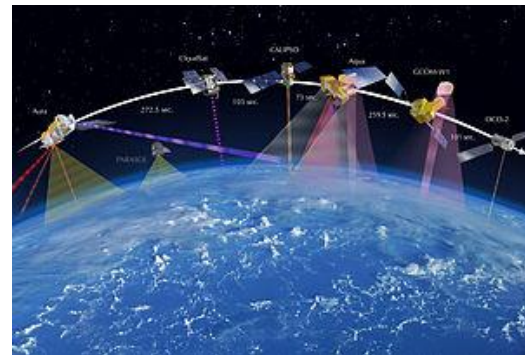
How to Solve Problems with It?

Three views of GIS

1. Smart, interactive, functional Map

2. A set of tools and procedures – A language to perform tasks

3. A well-managed system of information

DBMS

Data Files

ArcGIS: the Giant in GIS Software Industry

- **ArcGIS** is the overall name of the Esri Geographic Information Systems (GIS) software products.
- ArcGIS provides a standards-based platform for spatial analysis, data management, and mapping.

Non-Commercial GIS Software

- Open Source GIS - ***Desktop***
  - Quantum GIS (http://www.qgis.org/)
  - GRASS (https://grass.osgeo.org/)
  - MapWindow (http://www.mapwindow.org/)
- Open Source GIS - ***Server***
  - GeoServer (http://geoserver.org)
  - MapServer (http://mapserver.org/)
  - MapGuide Open Source (https://mapguide.osgeo.org/)
  - OpenLayers (http://openlayers.org/)
- Open Source GIS - ***Database Management System***
  - PostGIS (http://postgis.refractions.net/)
- Public GIS (free but not open source) - ***web map***
  - 2D (browser): Google Maps, Bing Maps, etc.
  - 3D (plugin): Google Earth, Virtual Earth, ArcGIS Explorer
- More on http://gislounge.com/open-source-gis-applications/
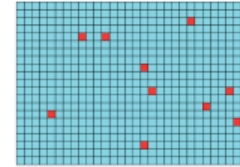
Basic GIS data formats
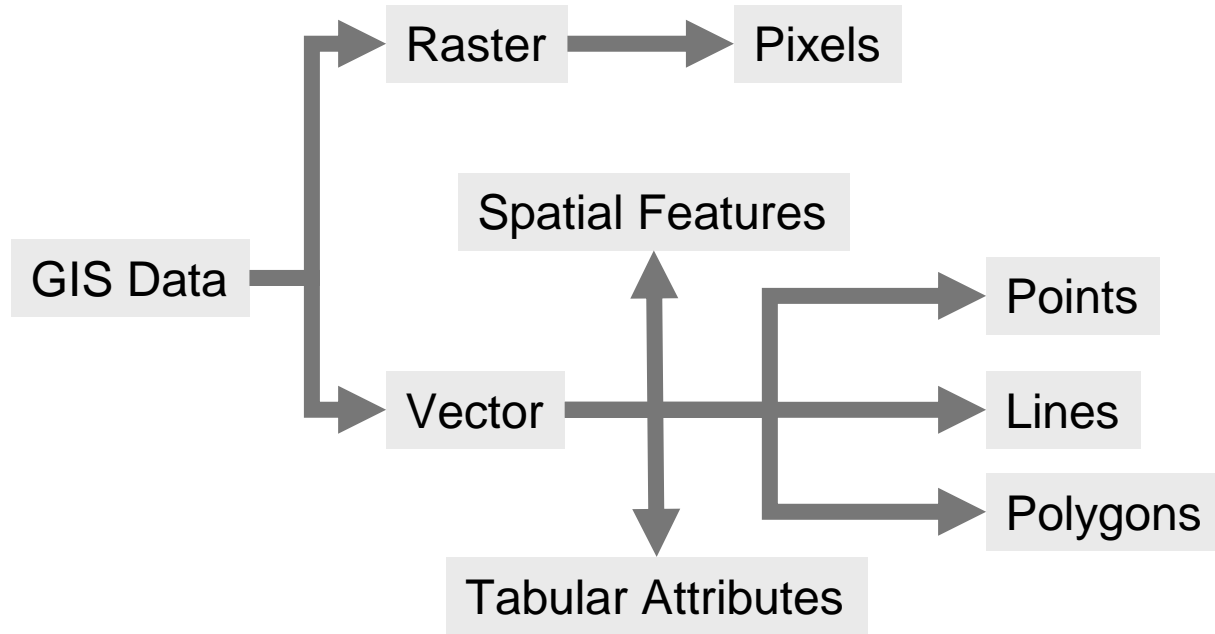
Raster Representation of Geographic Features

- **Point** - one pixel

- **Line** - a number of neighboring pixels strung out in a given direction

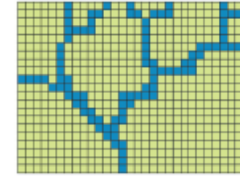- **Polygon** - an aggregation of neighboring pixels
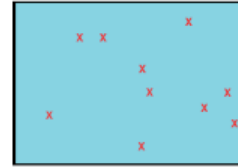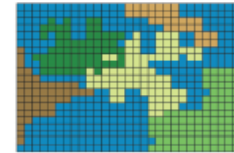
Point features

Raster point features

Polyline features
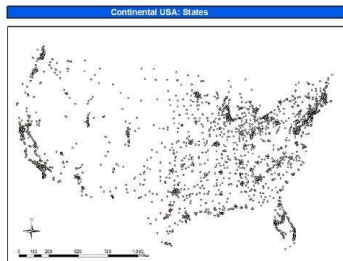
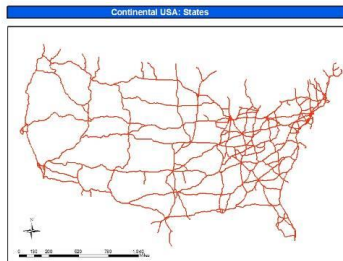Raster line features

Polygon features

Raster polygon features

Vector Representation of Geographic Features

- **Point**: a pair of x and y coordinates
- **Line**: a series of points connected together
- **Polygon**: a line enclosed from end to end

Examples:

Point: cities

Line: roads

Polygon: states

Uniqueness of GIS data

- **Location -** spatial registration of geographic features (***where on earth***).
- **Attribute -** tabular information associated with geographic features (***what, when***).
- **Topology -** spatial relationship among geographic features (***how***, *what is connected to what, what is next to what*).

✔ Scale
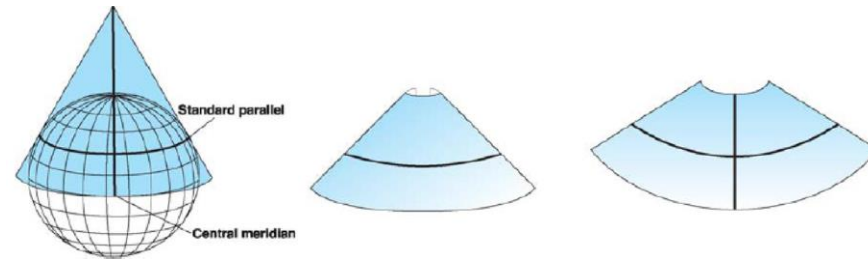✔ Resolution

✔ Projection
✔ Datum

Common Types of Map Projections
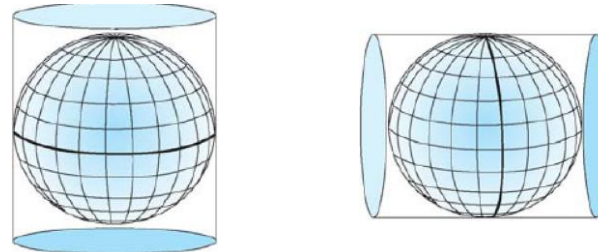
Planar Projections:
- Polar
- Equatorial
- Oblique

Conic Projections:
- Lambert Conformal

Cylindrical Projections:
- Mercator
- UTM

Polar   Equatorial   Oblique

Standard parallel

Central meridian

Major GIS Analysis Categories

Data Exploration
> Quantitative mapping
> Location based query
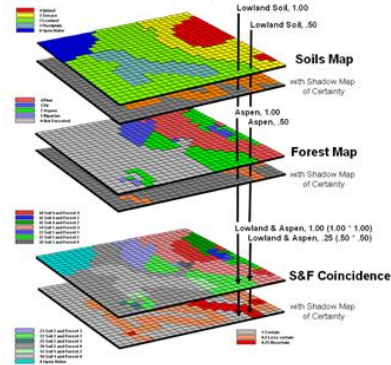
Data Analysis
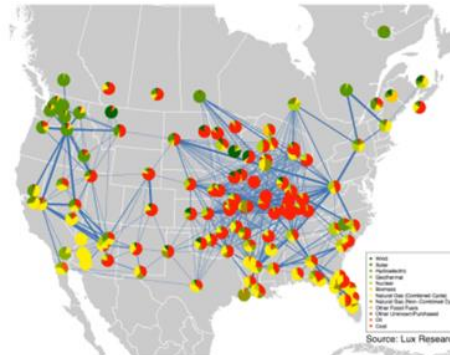> Overlay analysis
> Proximity analysis
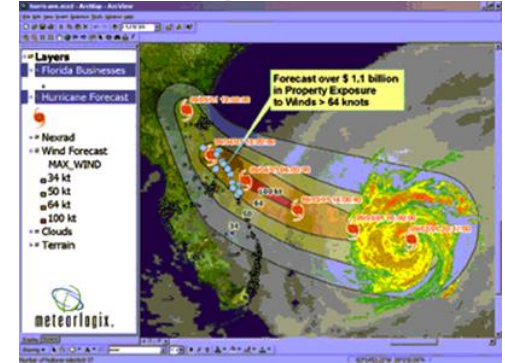> Network analysis
> Spatial statistical analysis
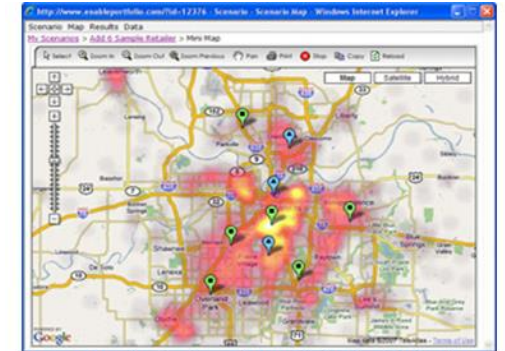> Geostatistical analysis

Habitat suitability analysis

Hurricane damage assessment


Lux Network Analysis of the US Electricity Grid

Hot spot analysis - Location optimization for business

GIS Analysis through Harvard Examples

- Visualizing Global Trade Dynamics in the Past Century (*Quantitative Mapping*)
- What is Available in Health Insurance Marketplaces (*Locate based Query*)
- Social-Economic Impacts of Fracking (*Overlay Analysis*)
- Tobacco Advertising Near Schools (*Proximity Analysis*)
- Selecting Hospital Sites by Ratio of Hospital to Diagnostic Center Distribution in Mexico City (*Network Analysis*)
- Exploring Attitude Towards Masks from Geo-tweet (*Spatial Statistics & Geostatistics*)
- The U.S. Cluster Mapping Project (*Comprehensive Analysis*)
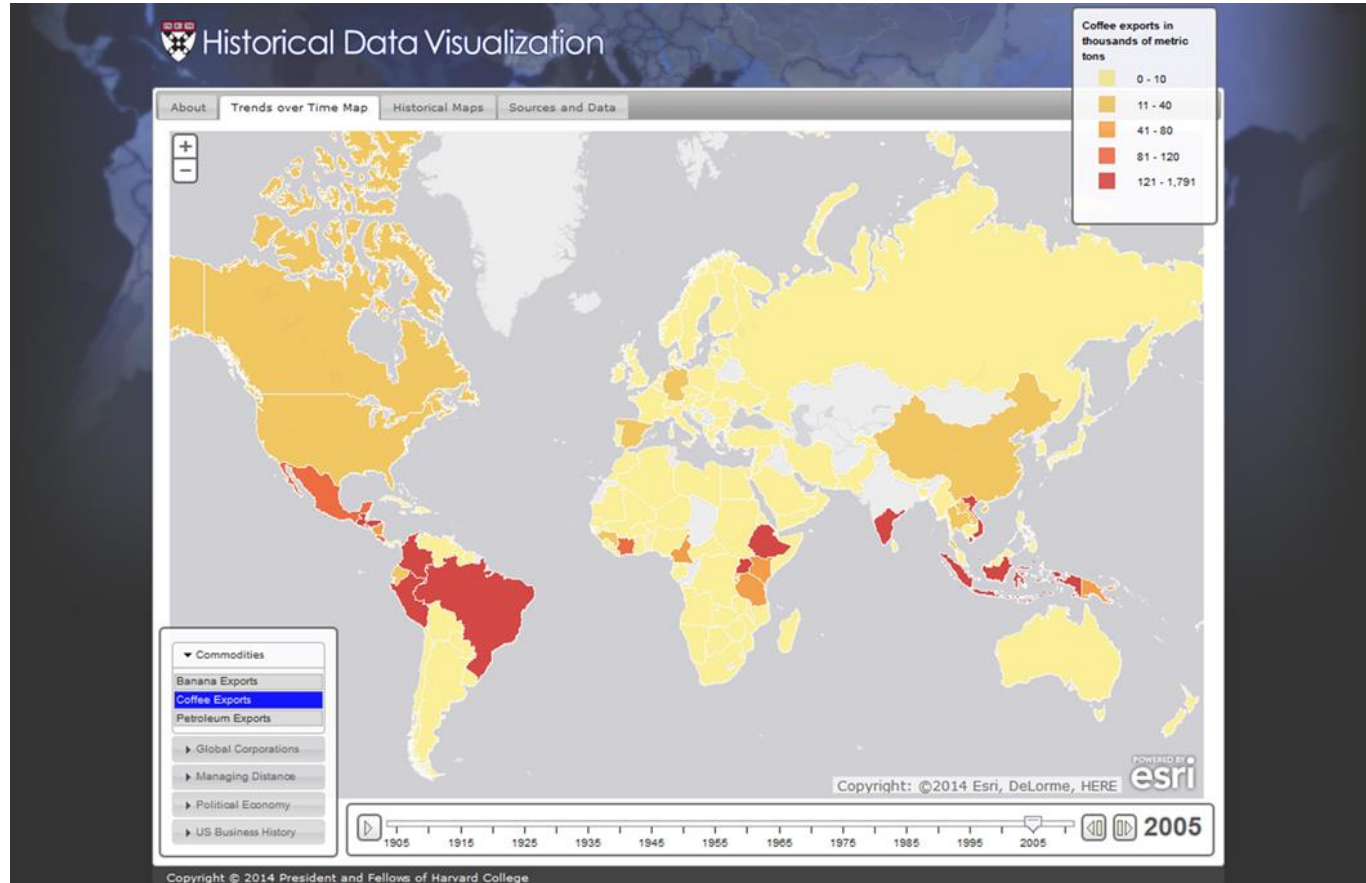
Quantitative Mapping: Classification by Attributes
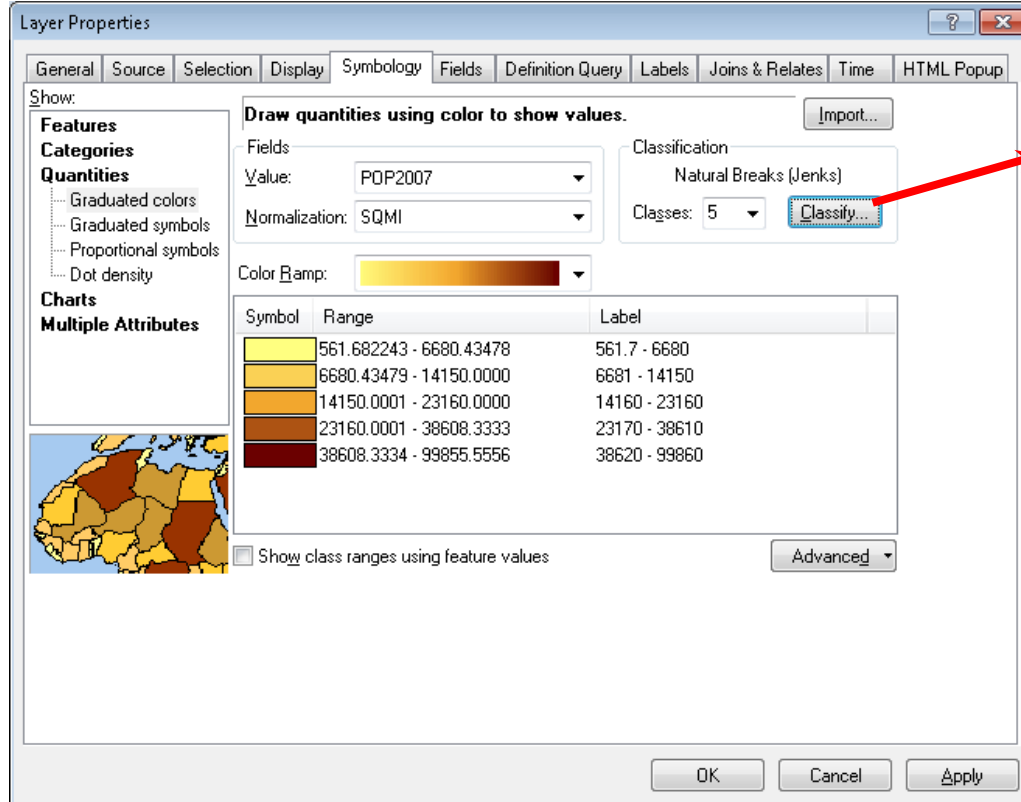

Classification Methods

- Manual
- Equal Interval
- Defined Interval
- Quantile
- Natural Breaks
- Geometrical Interval
- Standard Deviation

Quantitative Mapping: Visualizing Global Trade Dynamics in the Past Century

Location-based Query:
Select features by location in relationship with other features

- Identical to
- Intersect
- Within a distance of
- Have their centroid in
- Share a line segment with
- Crossed by the outline of
- Touch the boundary of
- Contain
- Contain (Clementini)
- Completely contain
- Within
- Within (Clementini)
- Completely within
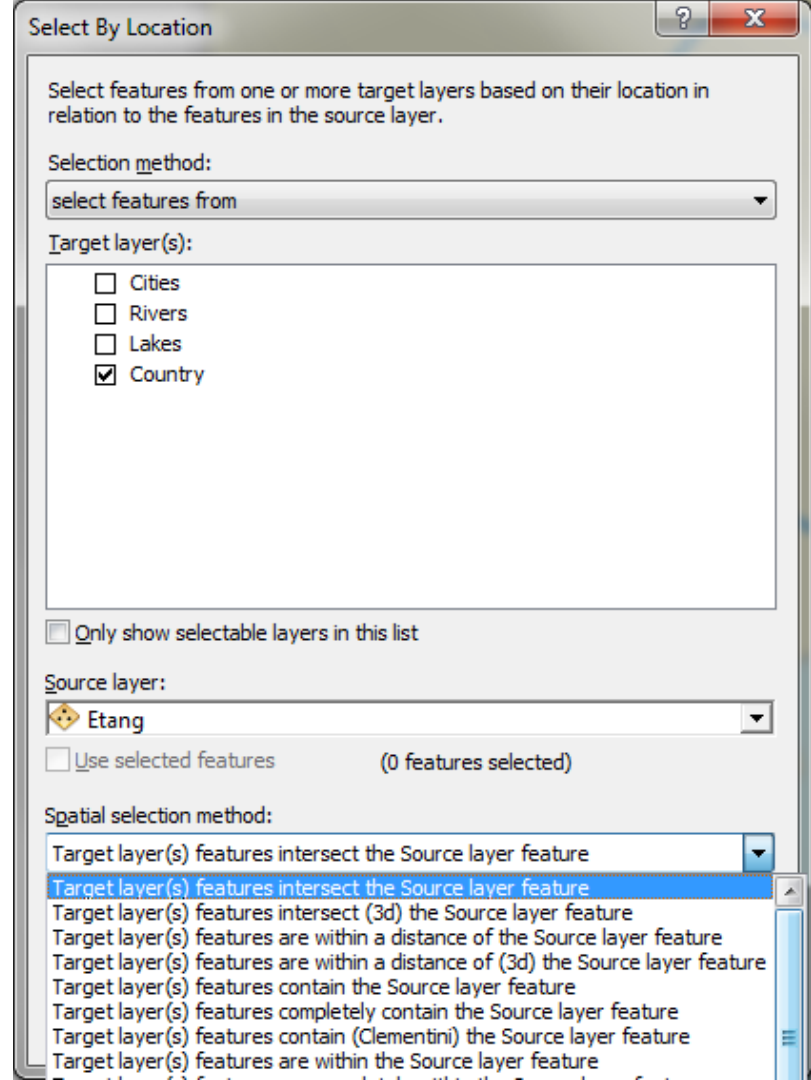
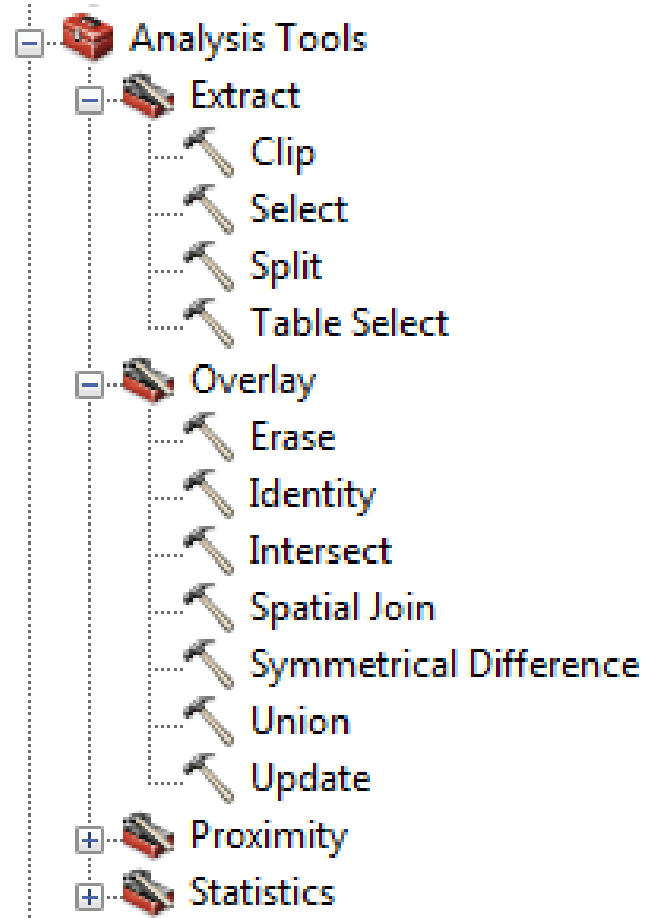# *Location Based Query:* What is Available in Health Insurance Marketplaces

The health insurance choices people have in the marketplaces depend not only on which state they live in but where within a state they live.
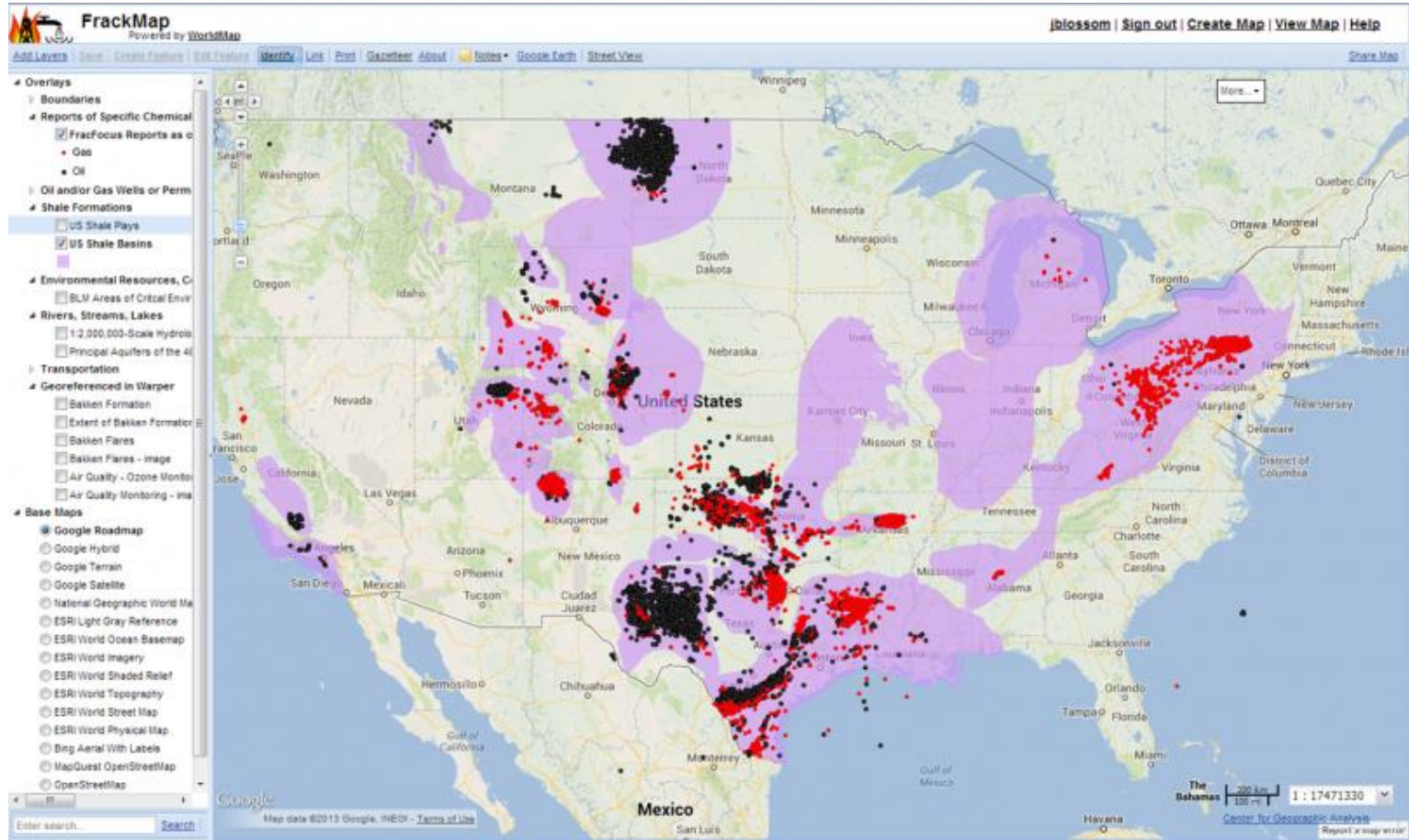
Overlay Analysis

Overlay operations by ArcGIS:
- Clip
- Split
- Erase
- Identity
- Intersect
- Symmetrical difference
- Union
- Update

Overlay Analysis: Social-Economic Impacts of Fracking

Proximity Analysis

- Also called **neighborhood** analysis, or **adjacency** analysis.
- Some location based query are proximity analysis, such as spatial joining.
- May involve one or multiple input feature classes.
- May produce a new feature class, or new attribute fields.
- Proximity operations in ArcGIS
  - Buffer and Multiple Ring Buffer
  - Create Thiessen Polygons
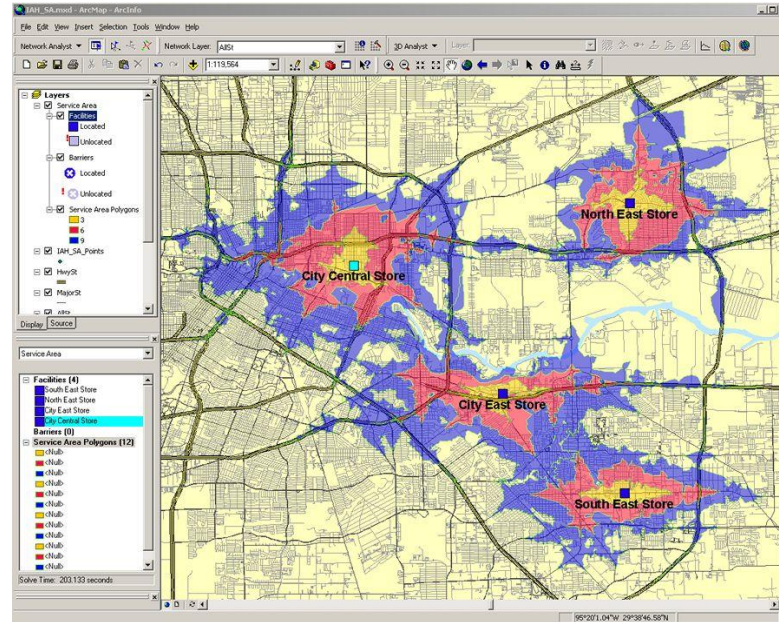 - Nearest
  - Point Distance
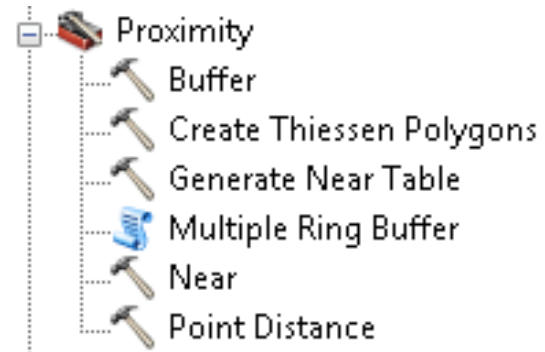
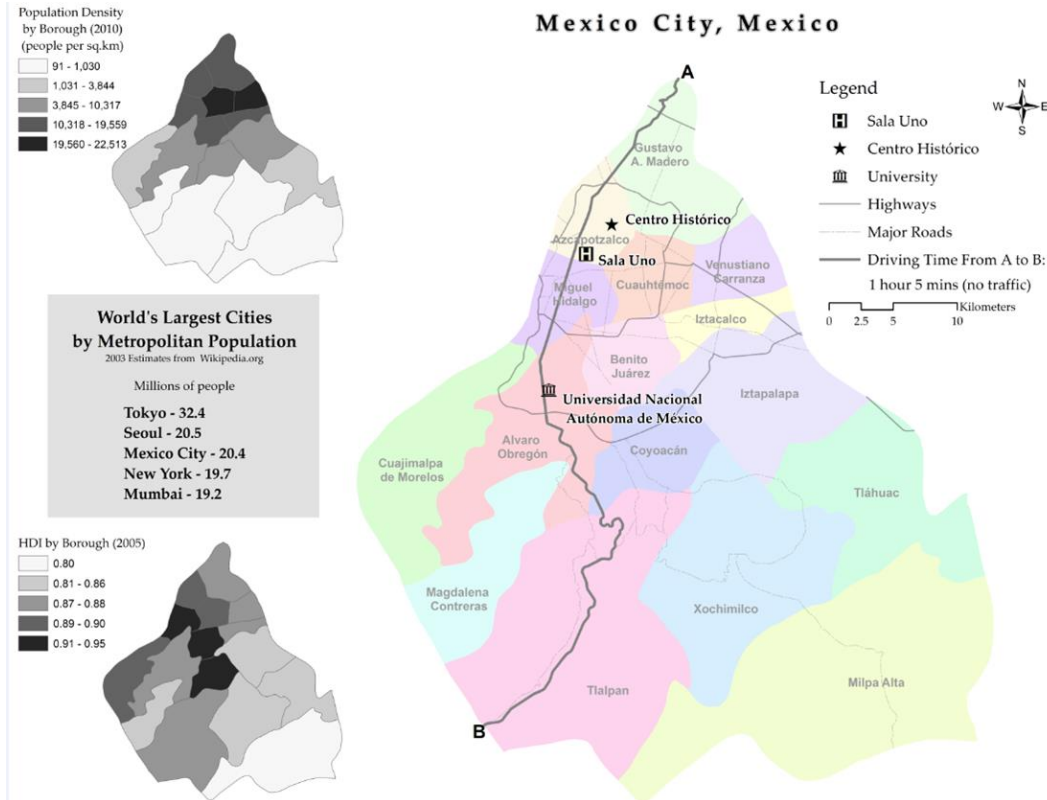# *Proximity Analysis:* Tobacco Advertising Near Schools
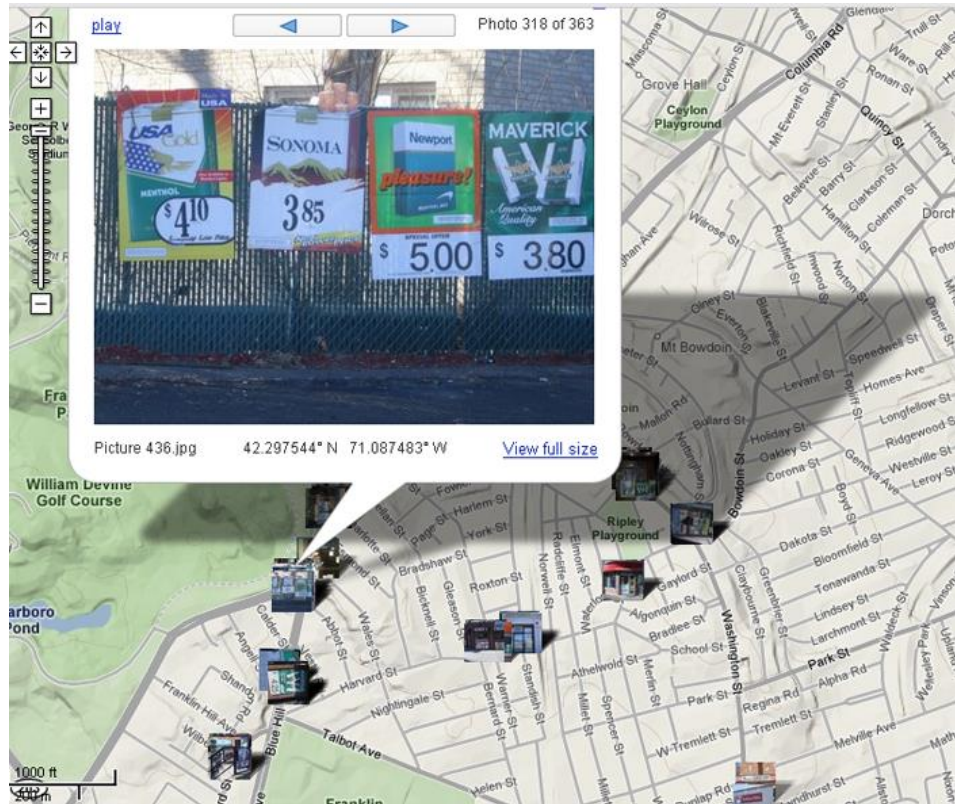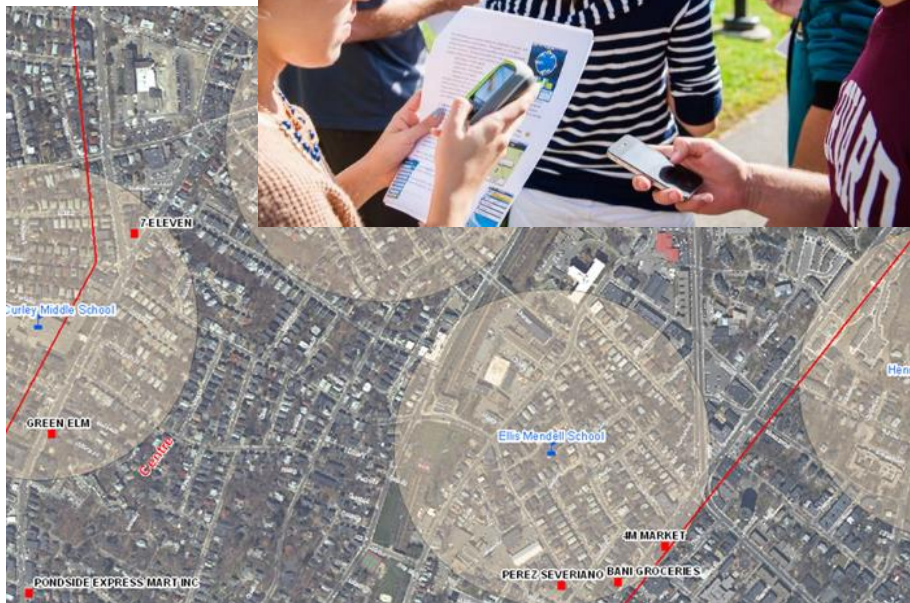
Network Analysis

- **Network Analysis** is a range of techniques for studying the properties of networks, including:

  - connectivity,

  - capacity, and

  - rates of flow.

Network Analysis: Selecting Hospital Sites by Ratio of Hospital to Diagnostic Center Distribution in Mexico City

This study is for a hospital company that performs cataract surgery. It is trying to select sites for placing hospitals that are **close** to the diagnostic centers.

Population Density
by Borough (2010)
(people per sq.km)
- 91 - 1,030
- 1,031 - 3,844
- 3,845 - 10,317
- 10,318 - 19,559
- 19,560 - 22,513

World's Largest Cities by Metropolitan Population
2003 Estimates from Wikipedia.org

Millions of people

Tokyo - 32.4
Seoul - 20.5
Mexico City - 20.4
New York - 19.7
Mumbai - 19.2

HDI by Borough (2005)
- 0.80
- 0.81 - 0.86
- 0.87 - 0.88
- 0.89 - 0.90
- 0.91 - 0.95

Mexico City, Mexico

Legend
- Sala Uno
- ★ Centro Histórico
- University
- Highways
- Major Roads
- Driving Time From A to B: 1 hour 5 mins (no traffic)

0 2.5 5   10 Kilometers

Gustavo A. Madero

★ Centro Histórico

Azcapotzalco

Sala Uno

Venustiano Carranza

Miguel Hidalgo

Cuauhtémoc

Iztacalco

Benito Juárez

Universidad Nacional Autónoma de México

Iztapalapa

Alvaro Obregón

Coyoacán

Cuajimalpa de Morelos

Tláhuac

Magdalena Contreras

Xochimilco

Milpa Alta

Tlalpan

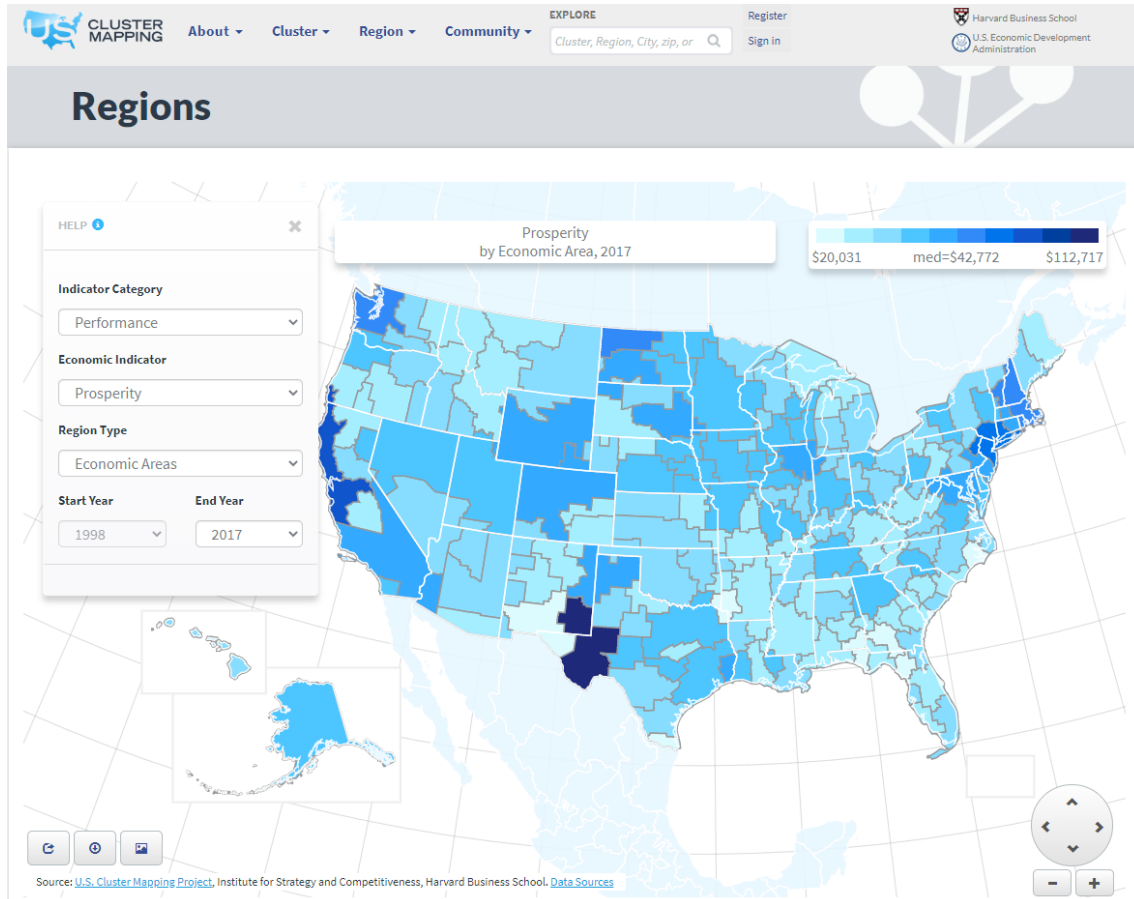# Spatial Statistics & Geostatistics

- It analyzes spatial distributions, patterns, processes, and relationships.
- It incorporates space (proximity, area, connectivity, and/or other spatial relationships) directly into the statistical mathematics.
    - Summarize spatial distributions (e.g. mean center, directional trend)
    - Identify statistically significant hot/cold spots or spatial outliers
    - Assess overall patterns of clustering or dispersion in space
    - Model spatial relationships
- It predicts values associated with spatial or spatiotemporal phenomena (e.g. interpolation, uncertainty)

Spatial Statistics: Exploring attitude towards masks from geo-tweet

Comprehensive Analysis: The U.S. Cluster Mapping Project

The U.S. Cluster Mapping Project is a national economic initiative that provides over 50 million open data records on industry clusters and regional business environments in the United States to promote economic growth and national competitiveness.

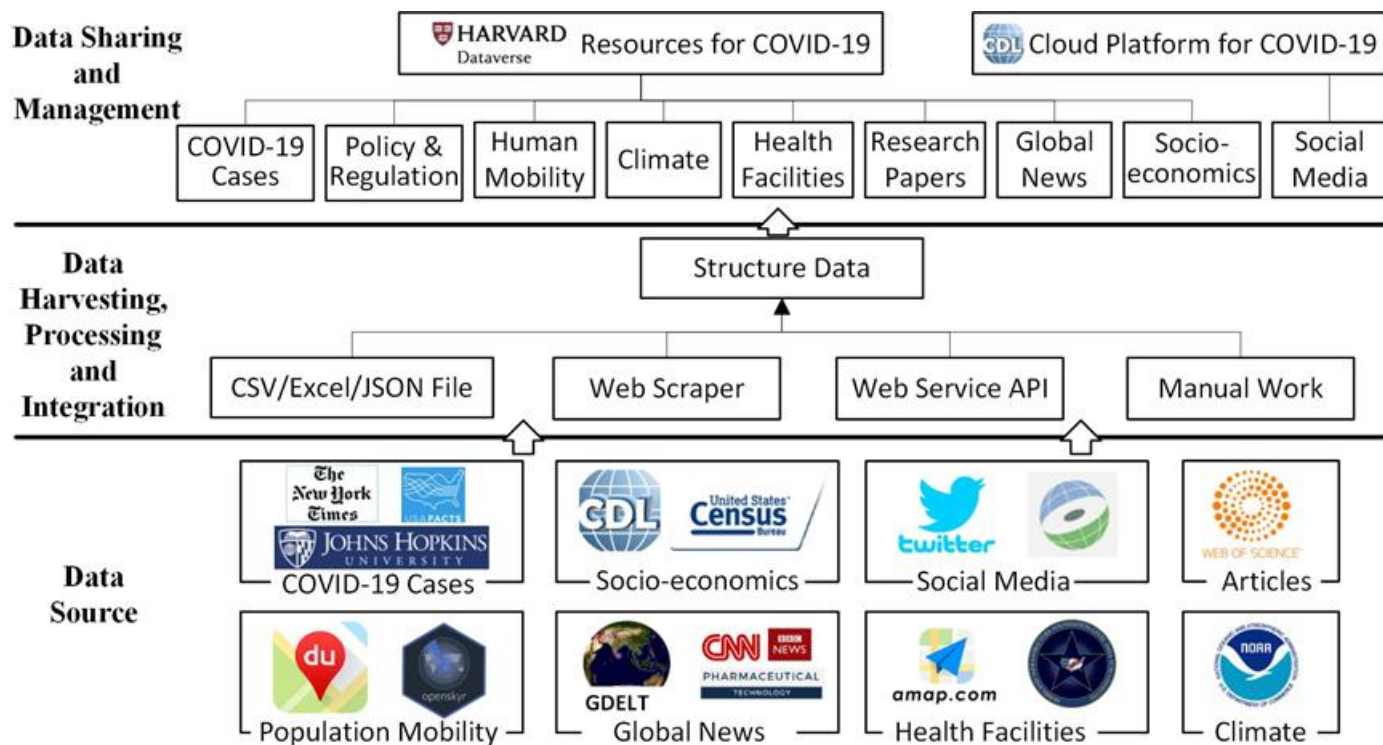# Supporting COVID-19 Research with GIS Resources

Economic	Social	Health
Unemployment 435,900+ new unemployment claims filed by King County residents from March 1 to July 11, 2020	**Food Insecurity** ↑ 17,300 more King County households received **Basic Food assistance** in June than January 2020, a 17.6% increase	**Healthcare** ↑ 10,000 more King County residents were enrolled in **Medicaid** in May than in January 2020
Housing 38% of calls to 2-1-1 hotline requested **housing-related assistance** from April 29-May 19, 2020	**Food Insecurity** ↑ 7% more women and children were enrolled in PHSKC's **WIC nutrition program** in April than January 2020	**Behavioral Health** ↑ 22% increase in calls to behavioral health crisis line in June 2020 compared to 2019
Traffic ↓ -24% decrease in local highway traffic on July 26, 2020 compared to 2019	**Domestic Violence** ↑ 16% increase in **domestic violence** calls to Seattle Police from Jan-April 2020 compared to 2019	**Mental Health and Suicide** 16% of middle and high school youth made a plan to attempt suicide in the past year before COVID-19
	Internet Access 67,000 King County households (7.5%) have **no internet access**	

- Fragmented information
- Incomparable and lack of standard
- Separate from maps
- Temporary data connection

↓

- Integration of data from different sources
- Standardization of data formats
- Integration of data and maps
- Permanent collections

https://kingcounty.gov/depts/health/covid-19/data/impacts.aspx

COVID-19 Data Collection, Integration and Sharing

Hu, T., Guan, W. W., Zhu, X., Shao, Y., Liu, L., Du, J., ... & Zhang, L. (2020). Building an Open Resources Repository for COVID-19 Research. *Data and Information Management*, *1*.

Data Collection and Integration

❑ Data standardization and association

❑ Rich spatial variables for geographic research
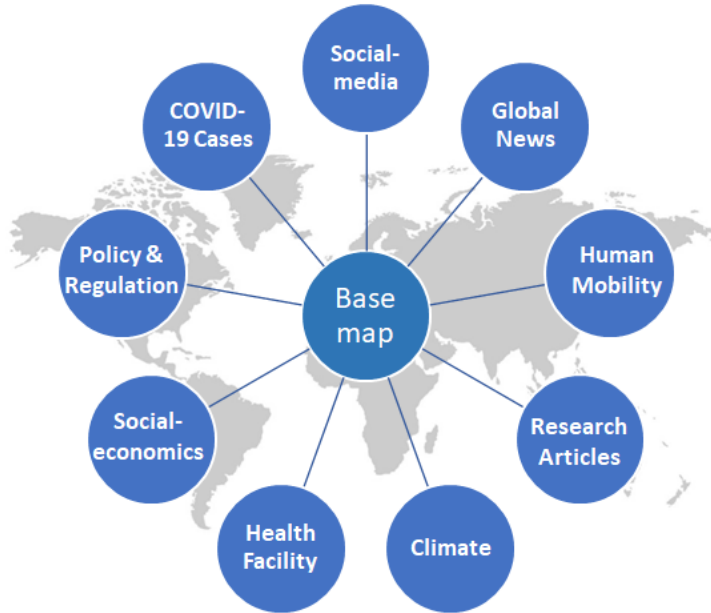
ID	Data Sets
1	Coronavirus cases
2	Population mobility
3	Health facilities
4	Traces
5	Flights
6	High-speed train
7	Global News
8	Social media
9	Policies
10	Meteorological data
11	Air Quality
12	Socioeconomic Data

Hu, T., Guan, W. W., Zhu, X., Shao, Y., Liu, L., Du, J., ... & Zhang, L. (2020). Building an Open Resources Repository for COVID-19 Research. *Data and Information Management*, *1*.

Data Sharing on Harvard Dataverse

Spatial Data Lab (SDL) for COVID-19 Research

- A **Data Library**

- An **Analytical Lab**

- A **Collaboration Hub**

- A **Training Center**

# Spatial Data Lab for COVID-19 Research
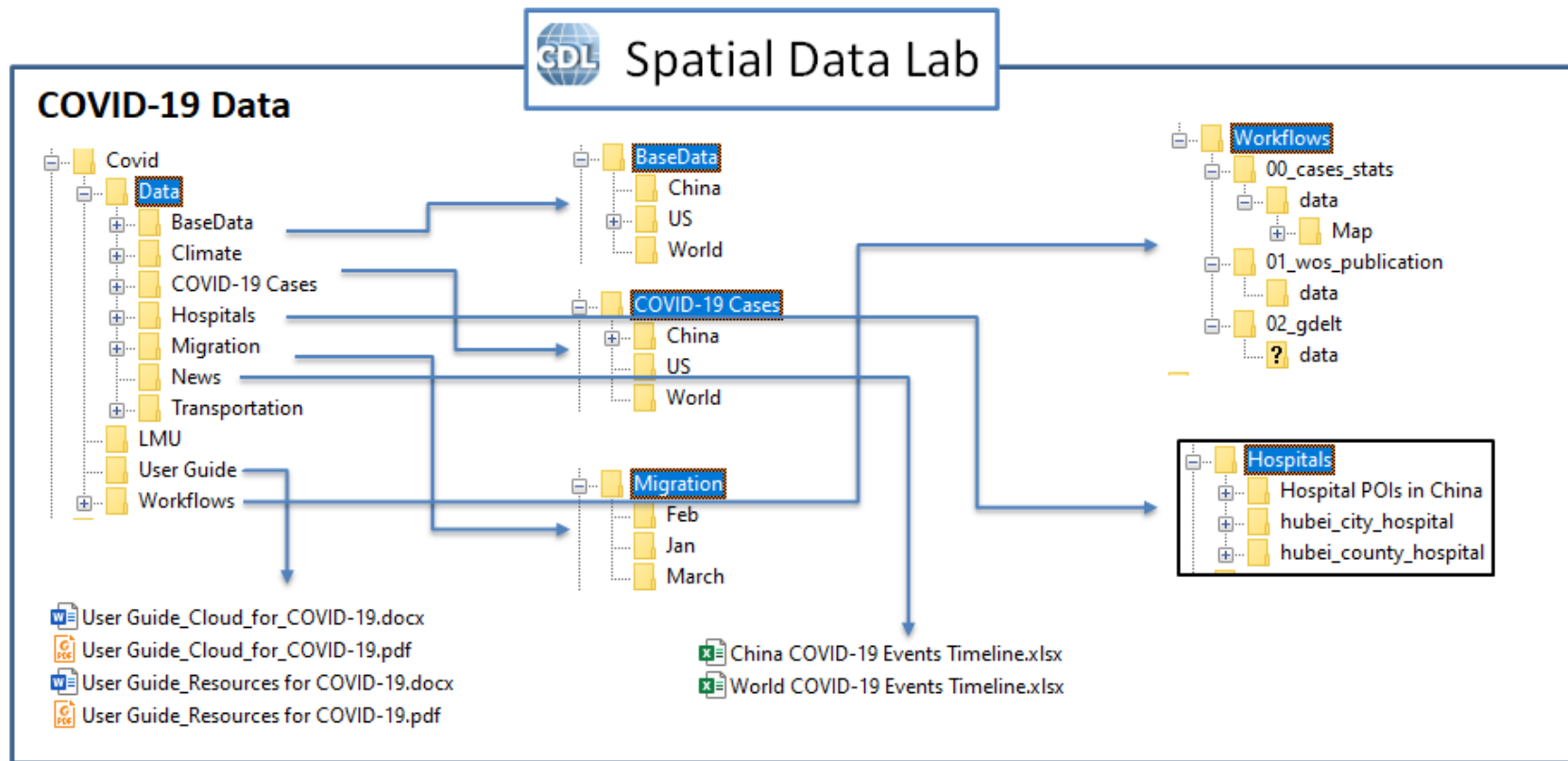
Data Sharing on Harvard Dataverse

Global users from over *150* countries & regions with over *250,000* data downloads

Accumulated Downloads of COVID-19 Resources at Harvard Dataverse

Global Access to "COVID-19 Resources" by Country

Global Distributions of Accessing "COVID-19 Resources" at Harvard Dataverse
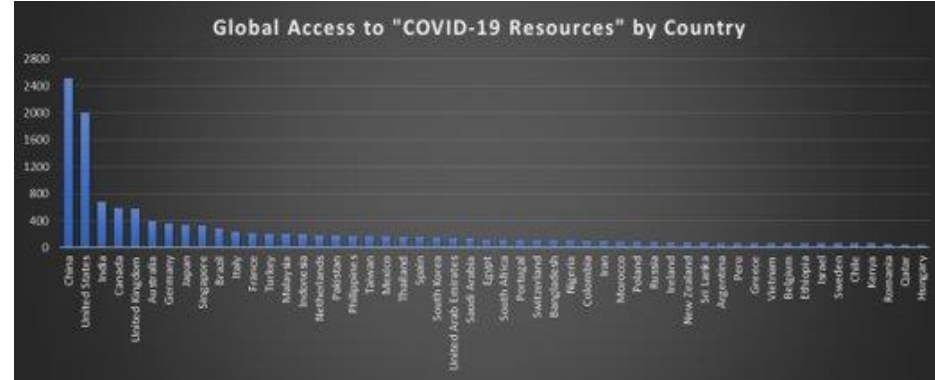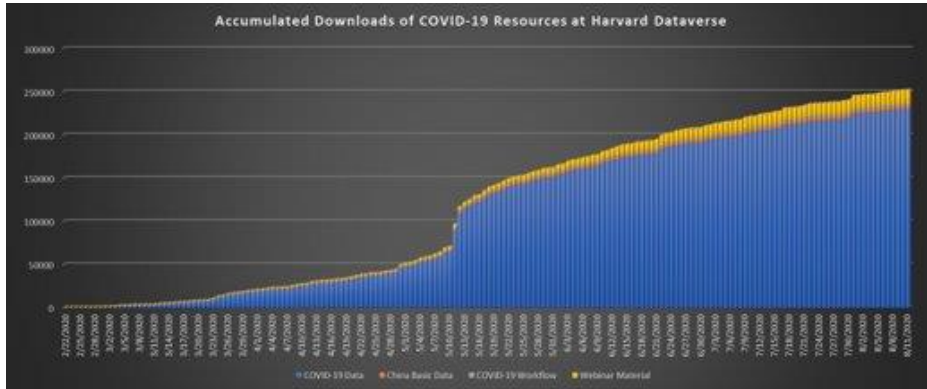
- Over *120* researchers from *9* countries applied for using the SDL platform for COVID-19 research, including *Oxford University*, *University of Illinois at Chicago*, *University of Maryland at College Park*, and so on.
- Research topics cover health inequality, economic loss, public opinions, correlation analysis between climate and virus spreading during COVID-19 epidemic, and so on.
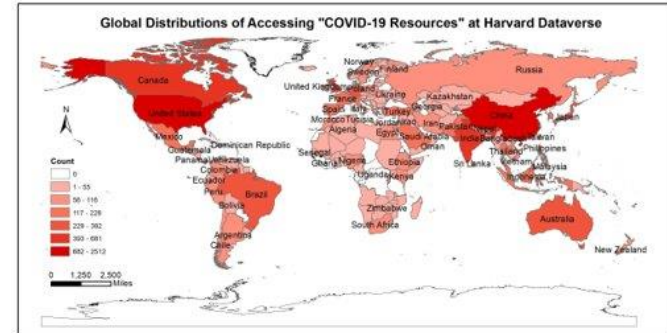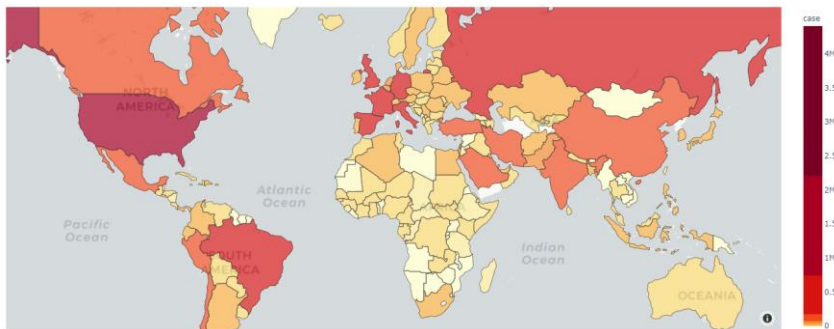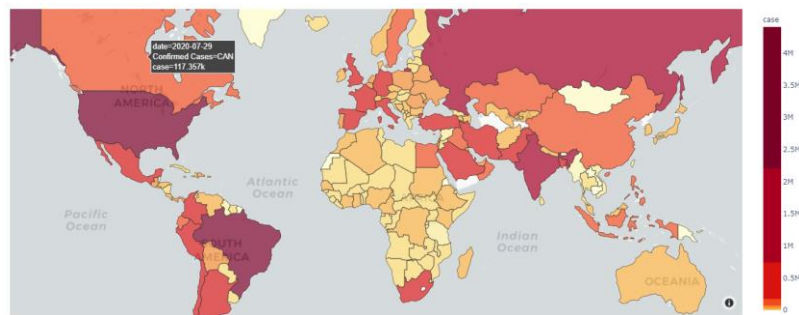
How did the virus spread worldwide?

01/29/2020
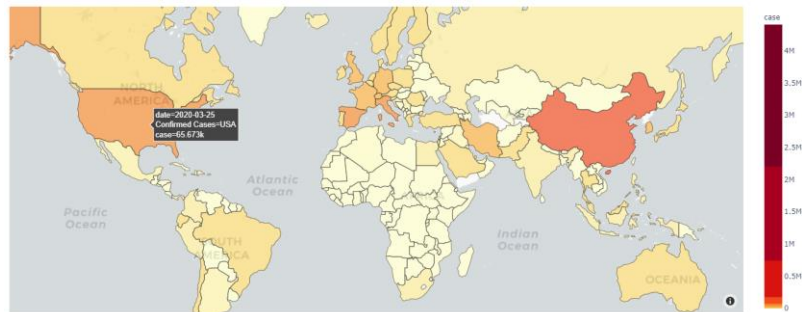
03/25/2020

05/27/2020

07/29/2020

How the global policies evolved and how the outbreak was (not) controlled?
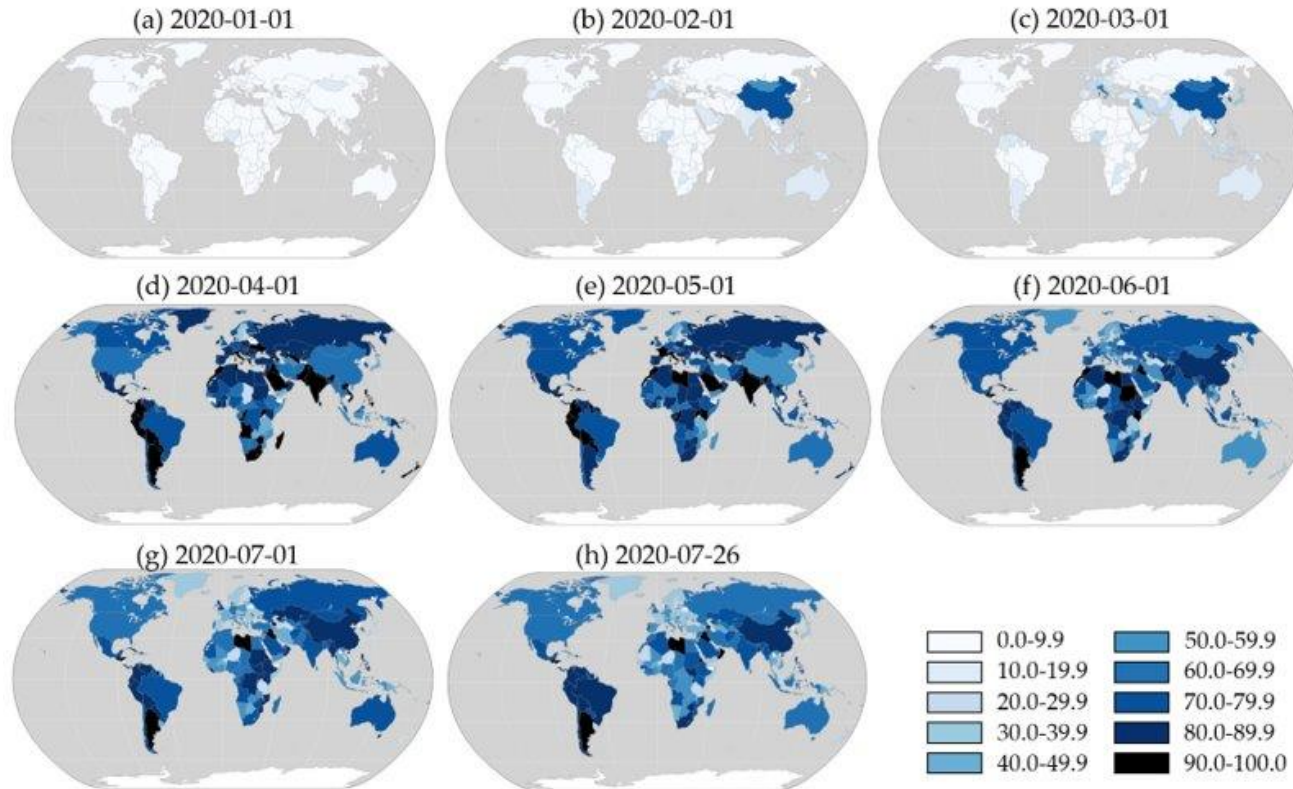
(a) 2020-01-01   (b) 2020-02-01   (c) 2020-03-01

(d) 2020-04-01   (e) 2020-05-01   (f) 2020-06-01

(g) 2020-07-01   (h) 2020-07-26

0.0-9.9	50.0-59.9
10.0-19.9	60.0-69.9
20.0-29.9	70.0-79.9
30.0-39.9	80.0-89.9
40.0-49.9	90.0-100.0

Yang, C., Sha, D., Liu, Q., Li, Y., Lan, H., Guan, W. W., ... & Wang, Z. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. *arXiv preprint arXiv:2005.04224.*

How are people movement changed in the U.S.?

Geotagged Tweets:

Population movements before and after the lockdown in the contiguous US.
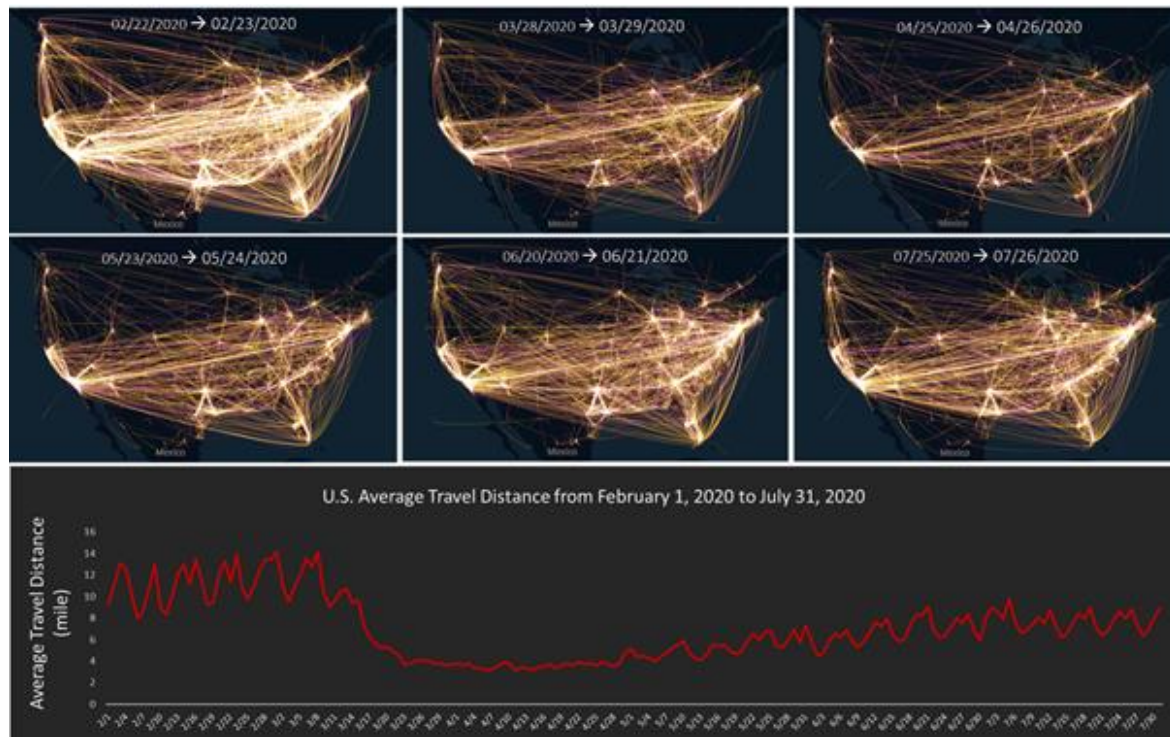
Yang, C., Sha, D., Liu, Q., Li, Y., Lan, H., Guan, W. W., ... & Wang, Z. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. *arXiv preprint arXiv:2005.04224*.

http://gis.cas.sc.edu/gibd/how-our-collective-efforts-of-fighting-the-virus-are-reflected-on-maps/

How are community activities changed?

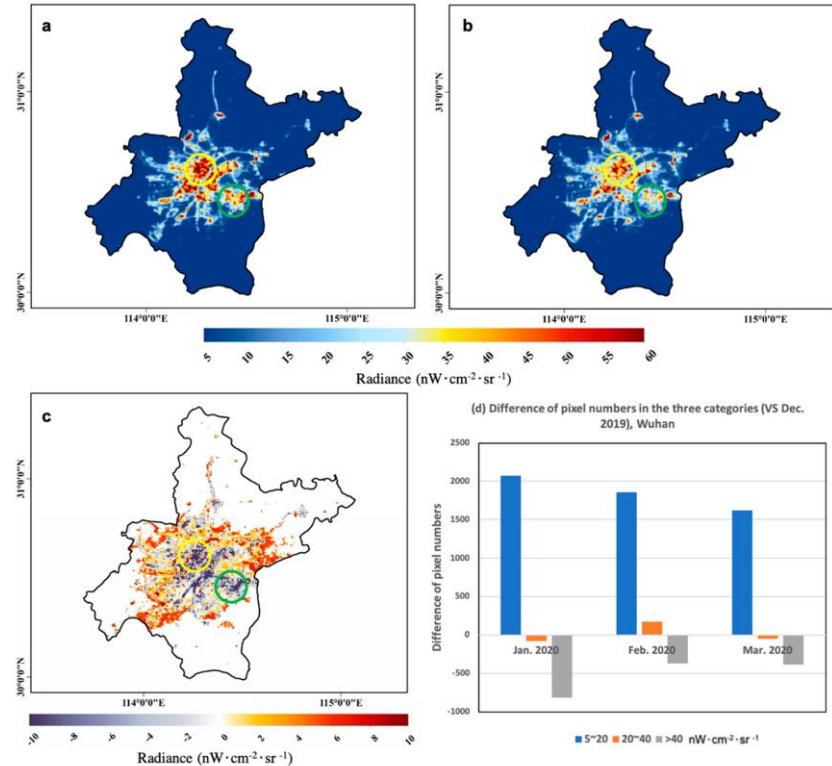**Google Community Mobility Report:**

community movement change during the outbreak of COVID-19 across different categories of places.

Yang, C., Sha, D., Liu, Q., Li, Y., Lan, H., Guan, W. W., ... & Wang, Z. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. *arXiv preprint arXiv:2005.04224*.

How people's daily lives are impacted by pandemic?

- these regions are dimmer during the lockdown, in February 2020, compared to December 2019
- There are significantly more pixels in the residential category and fewer pixels in the commercial centers than before

Liu, Q., Sha, D., Liu, W., Houser, P., Zhang, L., Hou, R., ... & Yang, C. (2020). Spatiotemporal Patterns of COVID-19 Impact on Human Activities and Environment in Mainland China Using Nighttime Light and Air Quality Data. *Remote Sensing*, *12*(10), 1576.

Summary: The Values of GIS

- It ***provides a framework*** for us to organize our knowledge – geographically.
- It ***reveals patterns, relationships and processes*** – broadens our knowledge and deepens our understanding.
- It helps us ***conceptualize, represent and communicate ideas*** – reconstruct history, illustrate the present, or plan for the future.
- Learning GIS is a ***life-long journey***.

ArcGIS Pro Functions

- Data creation and management
- Data exploration and visualization
- Cartography and mapping
- 3D rendering
- Spatial analytics and data science
- Image analysis
- Geoprocessing and Automation
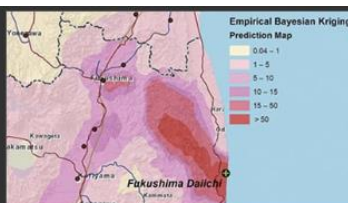- Sharing and publication
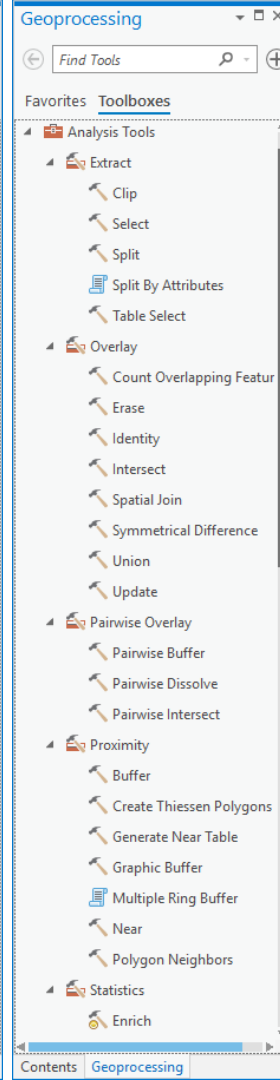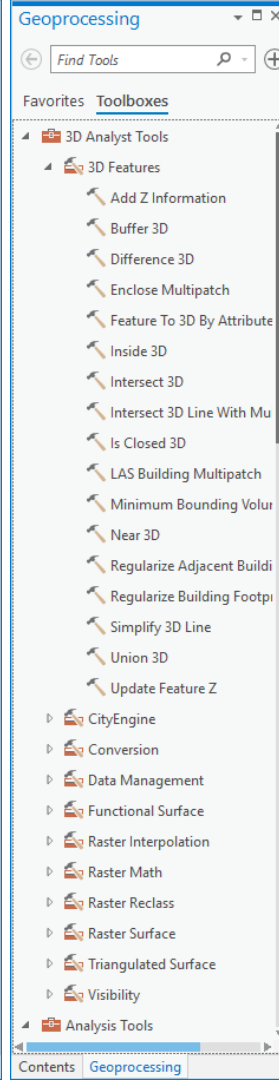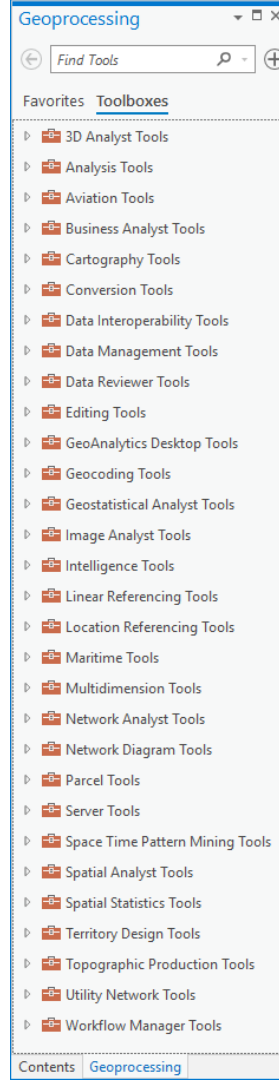
Major Extensions to ArcGIS Pro & ArcMap
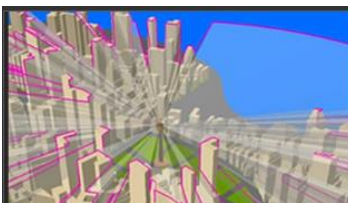
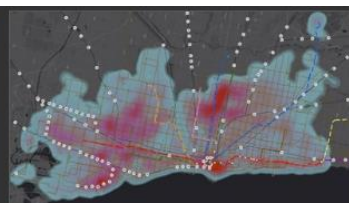**ArcGIS 3D Analyst**

Use a realistic perspective to analyze your data.

Learn more

ArcGIS Geostatistical Analyst

Investigate your data with advanced statistical tools.

Learn more

ArcGIS Network Analyst

Perform sophisticated routing and analysis on closest facilities or service areas.

Learn more

ArcGIS Spatial Analyst

Derive answers from your data using advanced spatial analysis.

Learn more

ArcGIS Image Analyst

Access a complete imagery analysis workstation.

Learn more

ArcGIS Business Analyst

Combine demographic, lifestyle, and spending data with map-based analytics to create accurate reports and present powerful market insights.

Learn more

GIS Tutorial Resources for Further Learning

- [Wrangling Data Into Maps](#)
- [Get started with ArcMap](#)
- [Spatial Analysis with ArcMap (5-part Video Tutorial)](#)
- [Get started with ArcGIS Pro](#)
- [ArcGIS Pro Resources](#)
- [Get started with ArcGIS Online](#)
- [QGIS Workshop and Video Tutorials](#)

Acknowledgment

- **Collaborators on COVID-19 Data Collection**
  - Wuhan University, Wuhan, China;
  - George Mason University, Fairfax, USA;
  - China Data Institute, Ann Arbor, US;
  - Shandong University of Science and Technology, Qingdao, China;
  - University of South Florida, Tampa, US;
  - University of Michigan, Ann Arbor, US;
  - East China Normal University, Shanghai, China;
  - Hefei University of Technology, Hefei, China;
- **Special thanks to the STCenter team at George Mason University (Professor Chaowei Phil Yang, Dexuan Sha, and Qian Liu), and Professor Zhenlong Li at University of South Carolina**
- **Harvard Dataverse Team**
- **Coronavirus Visualization Team (CVT)**

Thank you!

This work is partially funded by
NSF Awards #1841403 & #2027540

Contact us: https://gis.harvard.edu/contactus