Locational Big Data and Analytics: Implications for the Sharing Economy AMCIS 2017 SIGGIS Workshop

Brian N. Hilton, Ph.D.

Associate Professor

Director, Advanced GIS Lab

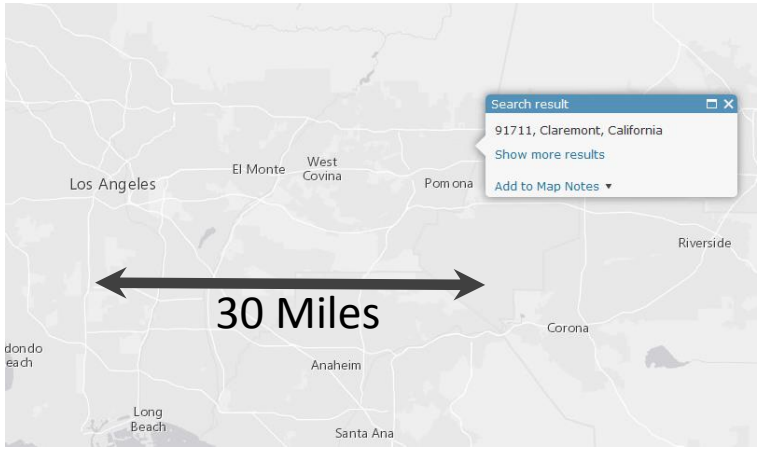
Center for Information Systems and Technology

Claremont Graduate University

Claremont, CA

Claremont Graduate University

CENTER FOR **INFORMATION SYSTEMS & TECHNOLOGY**



esri Development Center

Agenda

- Review of Three Research Projects / Use Cases
  - Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement
  - Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement
  - Stop-and-Frisk Policy from a Quantitative and Spatial Perspective
- Hands-On Demonstration (Sharing Economy Examples)
  - Insights for ArcGIS

Agenda

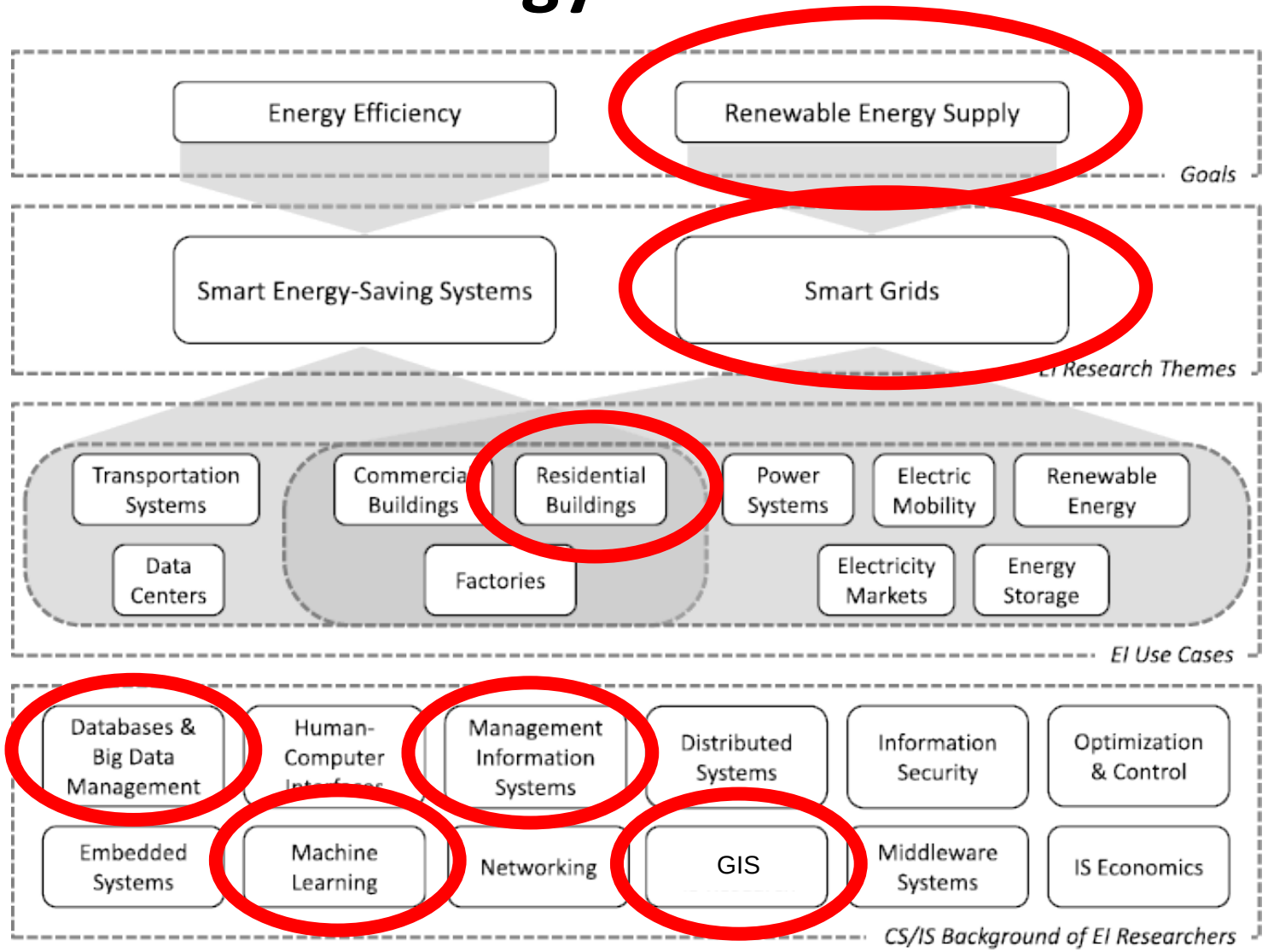
- Locational Big Data and Analytics has created a need for the efficient manipulation and scalable analysis of spatial big data on disparate, and distributed, datasets. As a result, this has opened a number of research areas such as:
  - Developing capabilities for accessing, formatting, and combining spatial big data in ways that enable it to be easily consumed;
  - Developing methodologies to derive insight into spatial big data for inferential understanding and decision making;
  - Developing teaching resources to better understand the use of data manipulation techniques, spatial statistics, and spatial data-mining tasks related to spatial big data; and
  - Developing novel spatial and spatiotemporal methods that can take advantage of newly emerging data-intensive computational resources.

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Domain – Energy Informatics

- Research Question – “Where are the optimal locations for the placement of Distributed Energy Resources, specifically, lithium-ion (Li-ion) batteries on the electricity grid?”

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement



Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Deploying **Distributed Energy Resources** in a widespread, efficient, and cost-effective manner requires complex integration with the existing electricity grid.
- The global scale-up of lithium-ion **(Li-ion) batteries** is enabling cost-effective energy storage systems for electric utility use.
- Policy incentives have increased **solar panel adoption** (grid-connected photovoltaic energy (PV) systems) – **California ranks first** among all states in number of solar PV systems installed.
- Research can identify and resolve the challenges of PV system integration, facilitating the transition to a smarter grid.

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

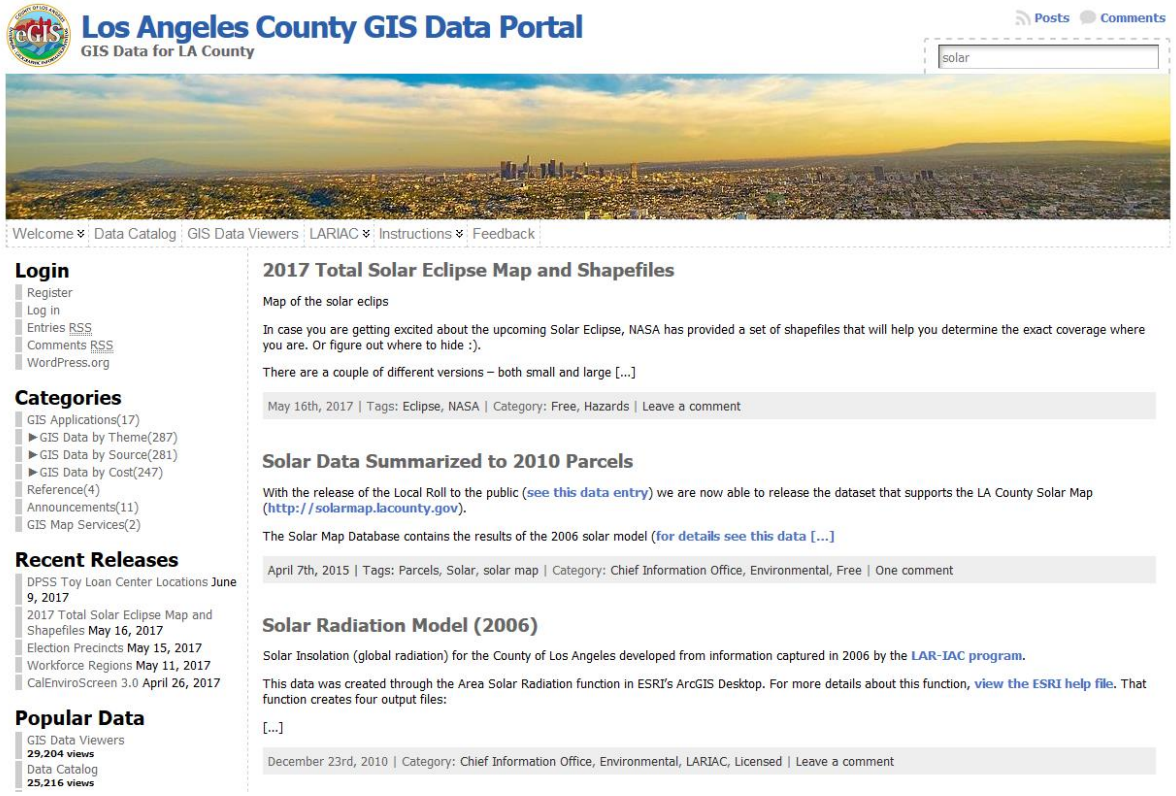
- Research steps:

1. Understanding **Solar Panel Adoption** across three main customer types: **Residential**, Commercial, and Industrial.

2. Development of GIS-based planning algorithm(s) for the optimal placement of new DERs (Li-ion batteries) given the spatial constraints of the existing electricity grid.

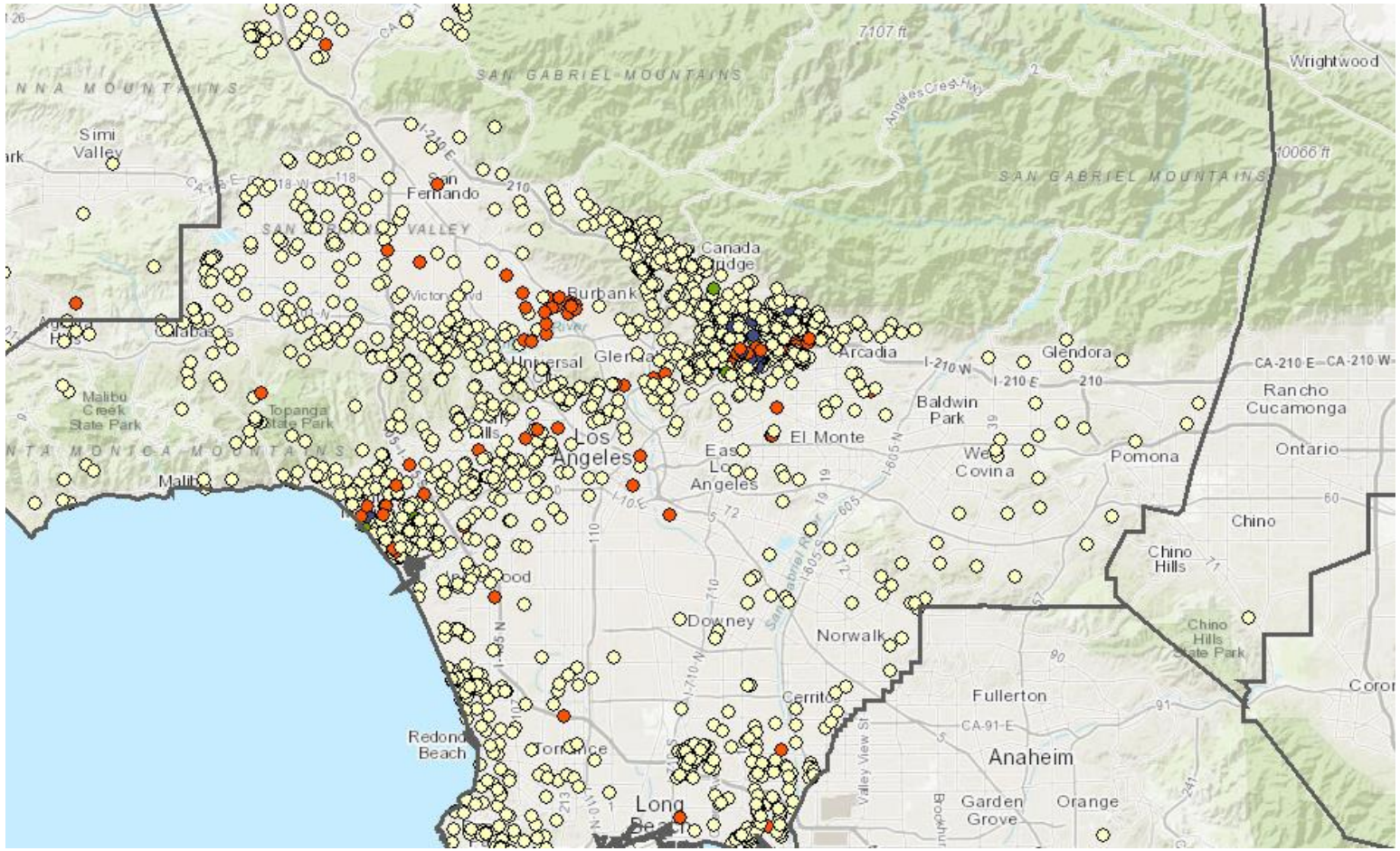
Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Data Provider
  - “The LA County GIS Data Portal is the place to search for GIS data created, maintained, licensed, and stored by the County of Los Angeles.”



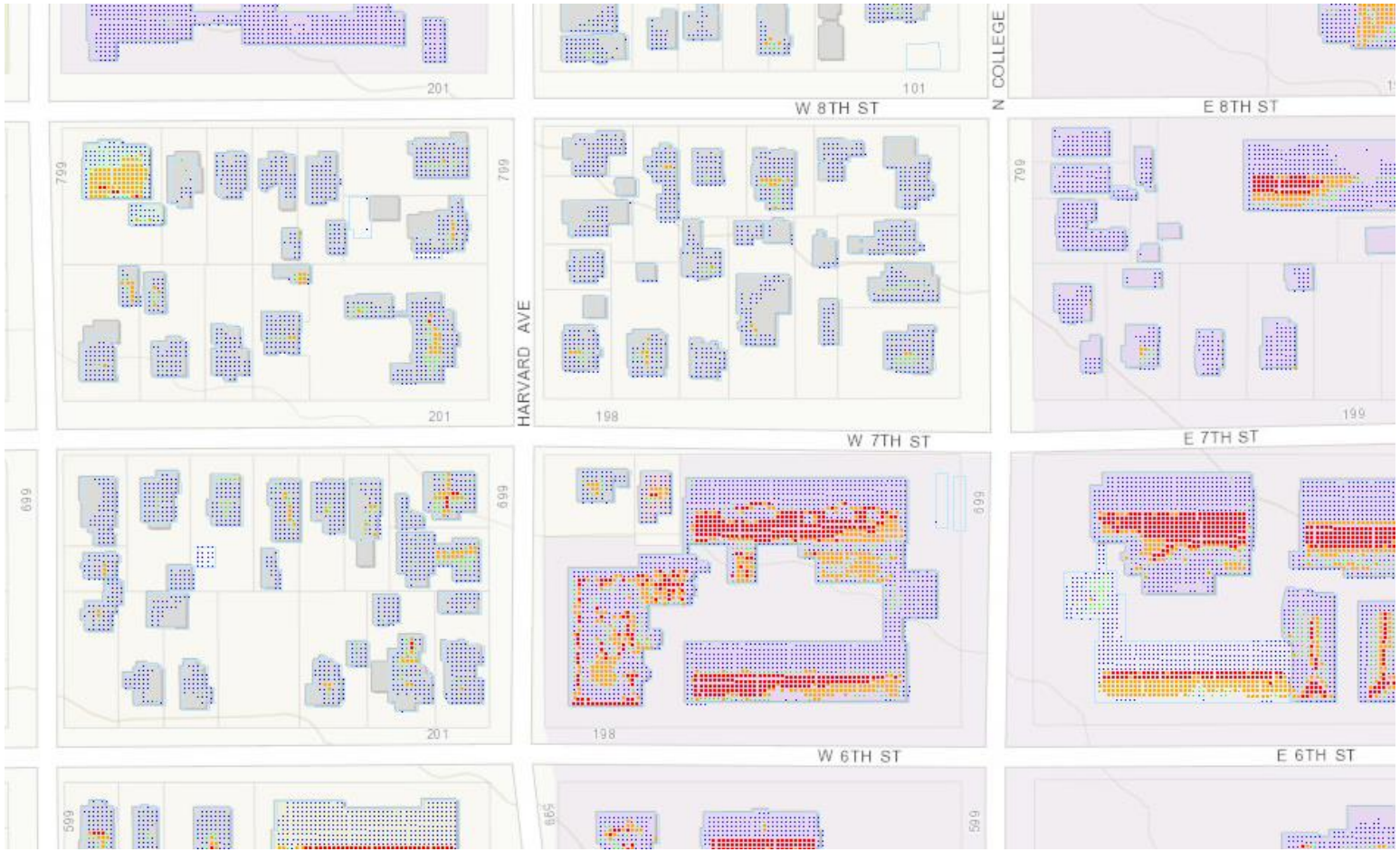
Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Solar Installation Data Description



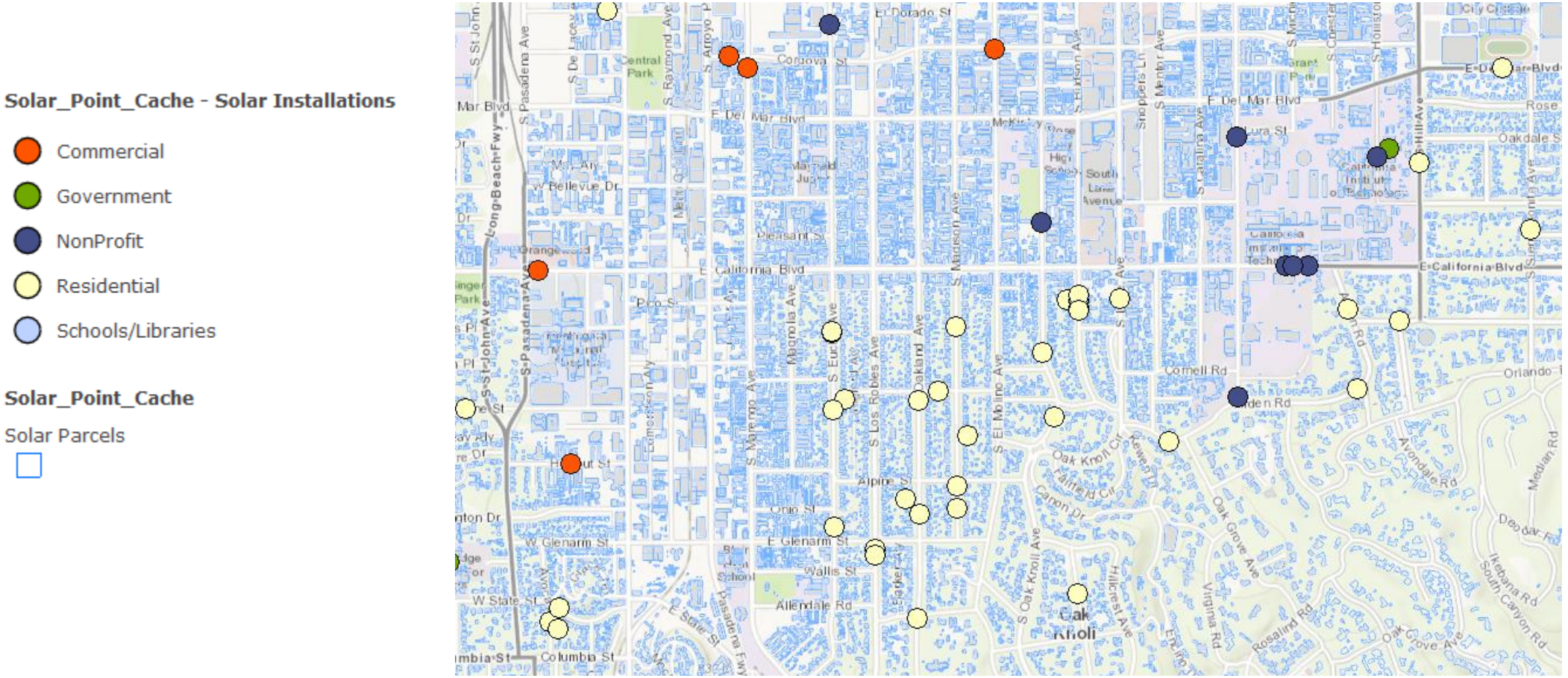
Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Solar PV Potential Data Description



Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Solar Installation and Solar PV Potential Data



Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Research Step (1)
- Creating a Predictive Model for Residential Solar Panel Adoption
  - *Los Angeles County: Residential Parcels = 1,868,519 out of 2,392,100 (78%)*

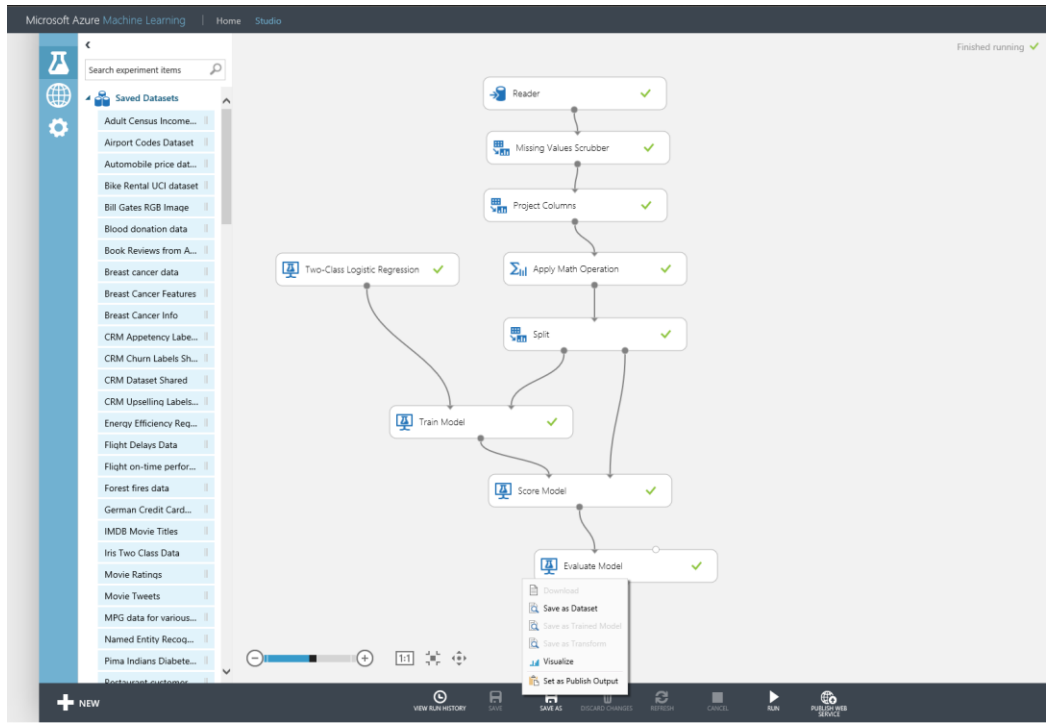
    - Dependent variable:
      Likelihood of a household adopting solar energy panels

    - Independent variables:
      (1) parcel information, such as: parcel age, parcel value, etc.
      (2) customer demographics, such as: household income, household size, etc.
      (3) expenditure data, such as: electricity usage, mortgage value, etc.

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Research Step (1)
- The results of the Two-Class Logistic Regression (Azure Machine Learning) indicated that six factors emerged as significant predictors of solar adoption:

  - parcel age,
  - average household size,
  - total area suitable for solar roof top,
  - total building area square feet,
  - average household income, and
  - average home value

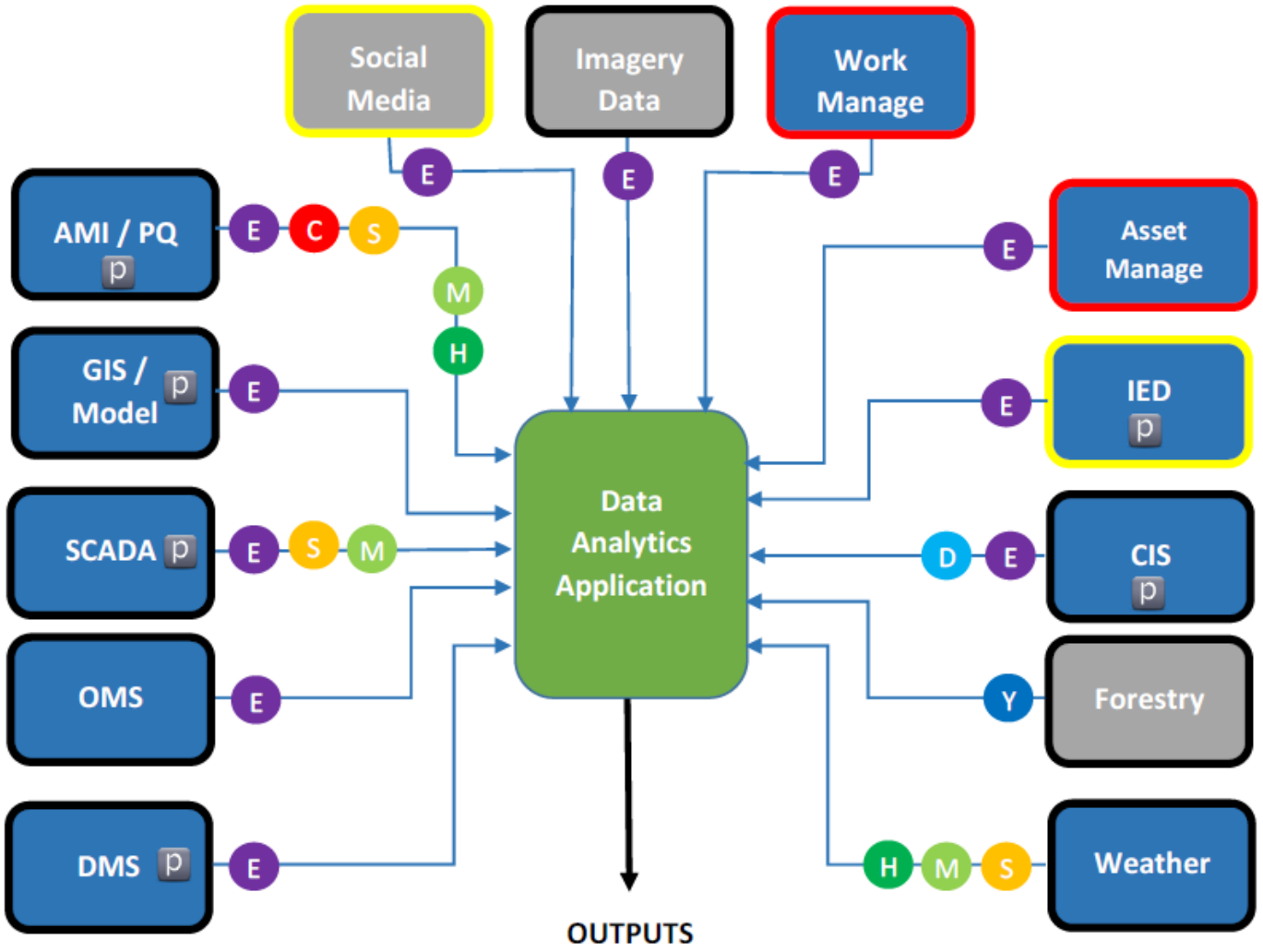


Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

- Future Work / Analysis
  - Research Step (2)
    - Development of GIS-based planning algorithm(s) for the optimal placement of new DERs
  - Data Provider
    - “The Electric Power Research Institute, or EPRI, conducts research on issues related to the electric power industry.”

- Devise a methodology for organizing our disparate datasets…

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement



Classifications of Data:

- Traditional Data Set
- New Data Set
- Structured Data
- Un-structured Data
- Format of Data Varies

(p) Denotes a primary data set used to solve this data analytics case.

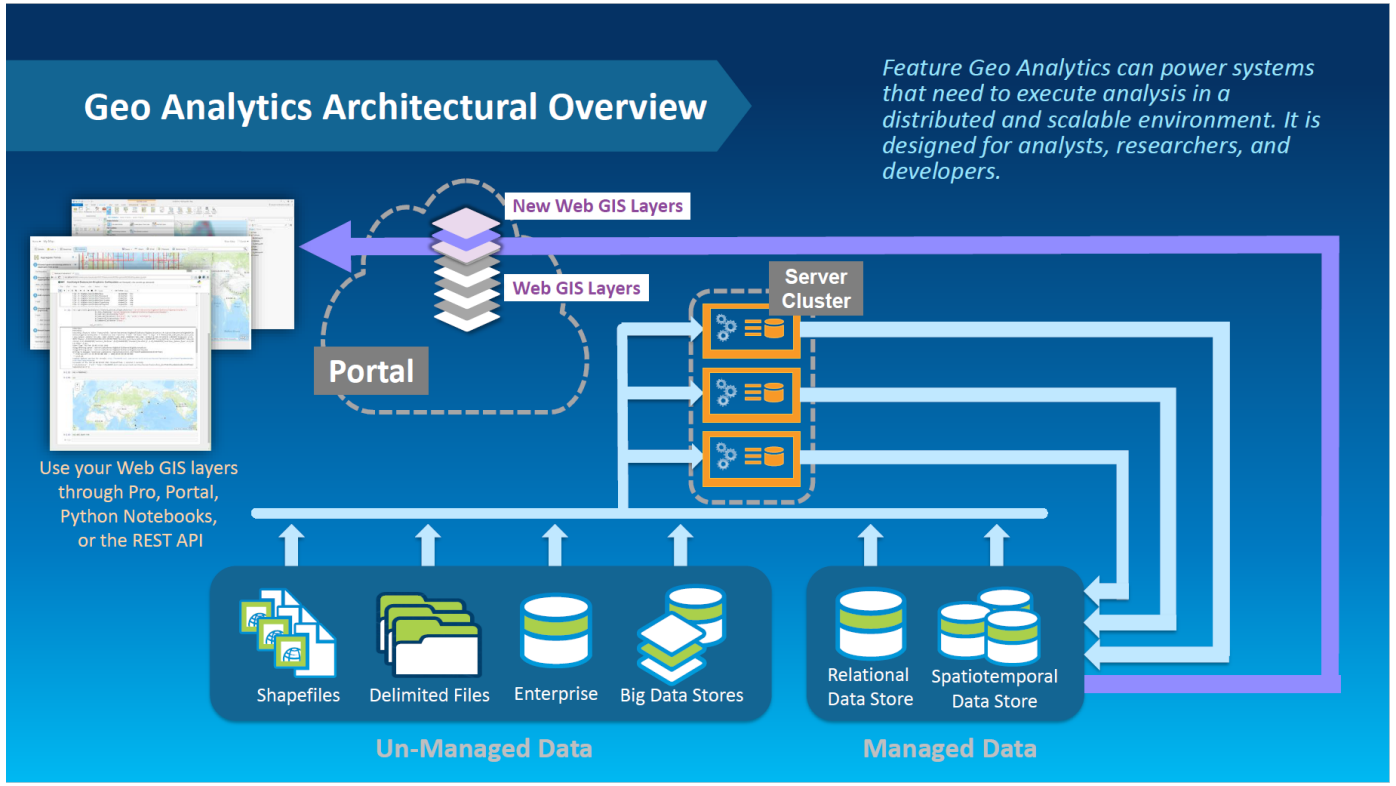
Frequency of Measurement

- C — Cycles
- S — Seconds
- M — Minutes
- H — Hours
- D — Days
- Y — Months to Years
- E — Event Driven

Optimization Algorithm for Spatially Constrained Distributed Energy Resource Placement

Developing capabilities for accessing, formatting, and combining spatial big data in ways that enable it to be easily consumed.

- ArcGIS GeoAnalytics Server



Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- Domain – Logistics / Freight Network Planning

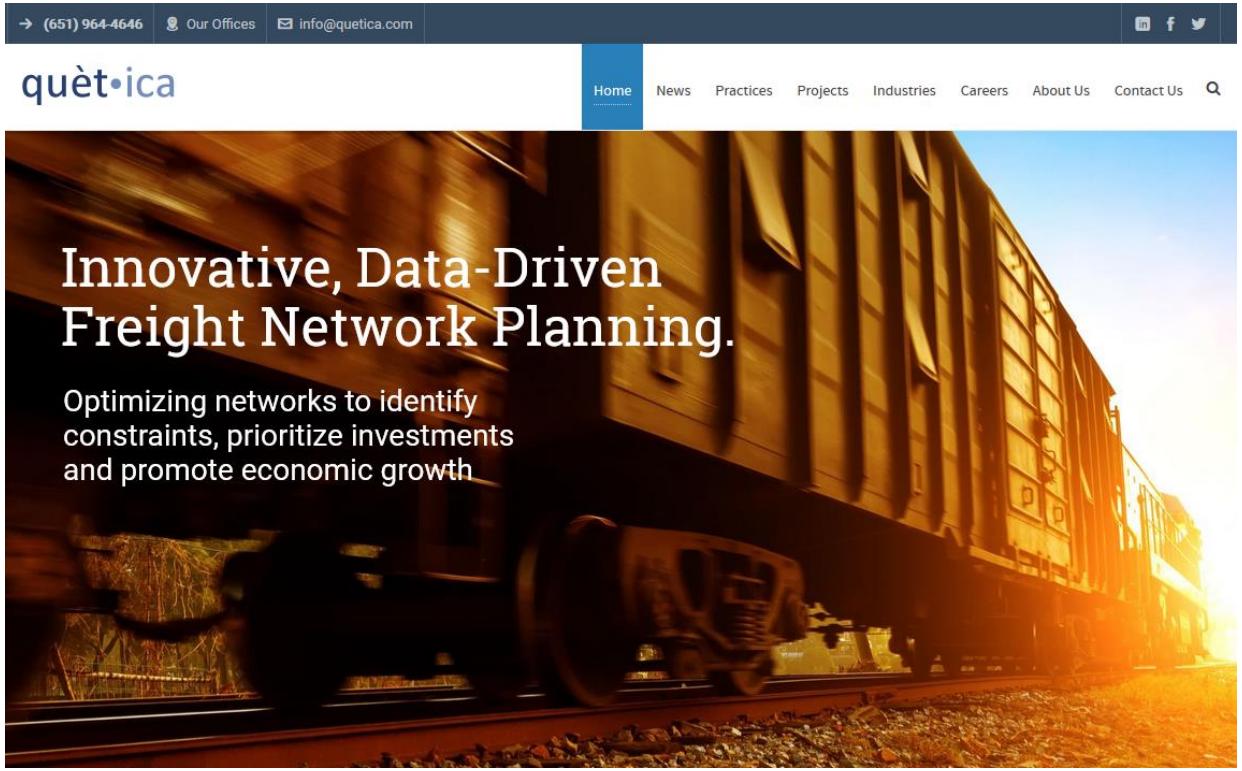
- Research Question – “How can we better understand commodity flows for economic development, for freight policy analysis, and transportation infrastructure impacts?”

Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- In Minnesota, technological and economic shifts in the grain supply chain have altered the way grain producers and sellers navigate their local freight network.
- In particular, many producers have been increasing their personal trucking capacity and taking longer trips to intermodal and domestic market options.
- This logistical reshaping of local grain supply chains pressure transportation officials to reconsider the consequences for road infrastructure and congested freight corridors.

Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- Data Provider
  - Quetica, a Minnesota-based, supply chain management company that uses commodity flow analysis to optimize freight network planning.



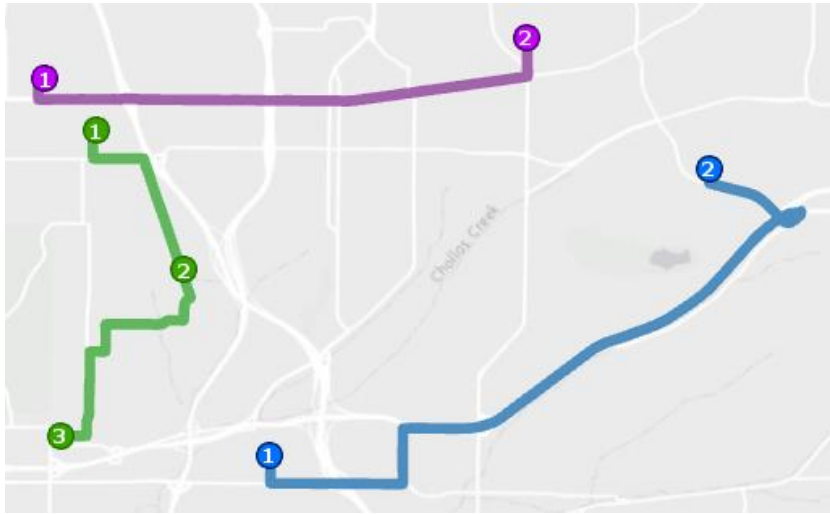
Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- Freight Data Description
  - The Quetica sample dataset included **cereal grain shipments, via truck**, including **shipment weight**, for Midwest **U.S. counties** in 2014:

  - 257,006 - Midwest U.S. shipments - total tons 764,848,291
  - 15,920 - MN-related (internal/external) shipments - total tons 79,638,868
  - 4,489 - MN-only shipments (internal/internal) - total tons 66,789,589

Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

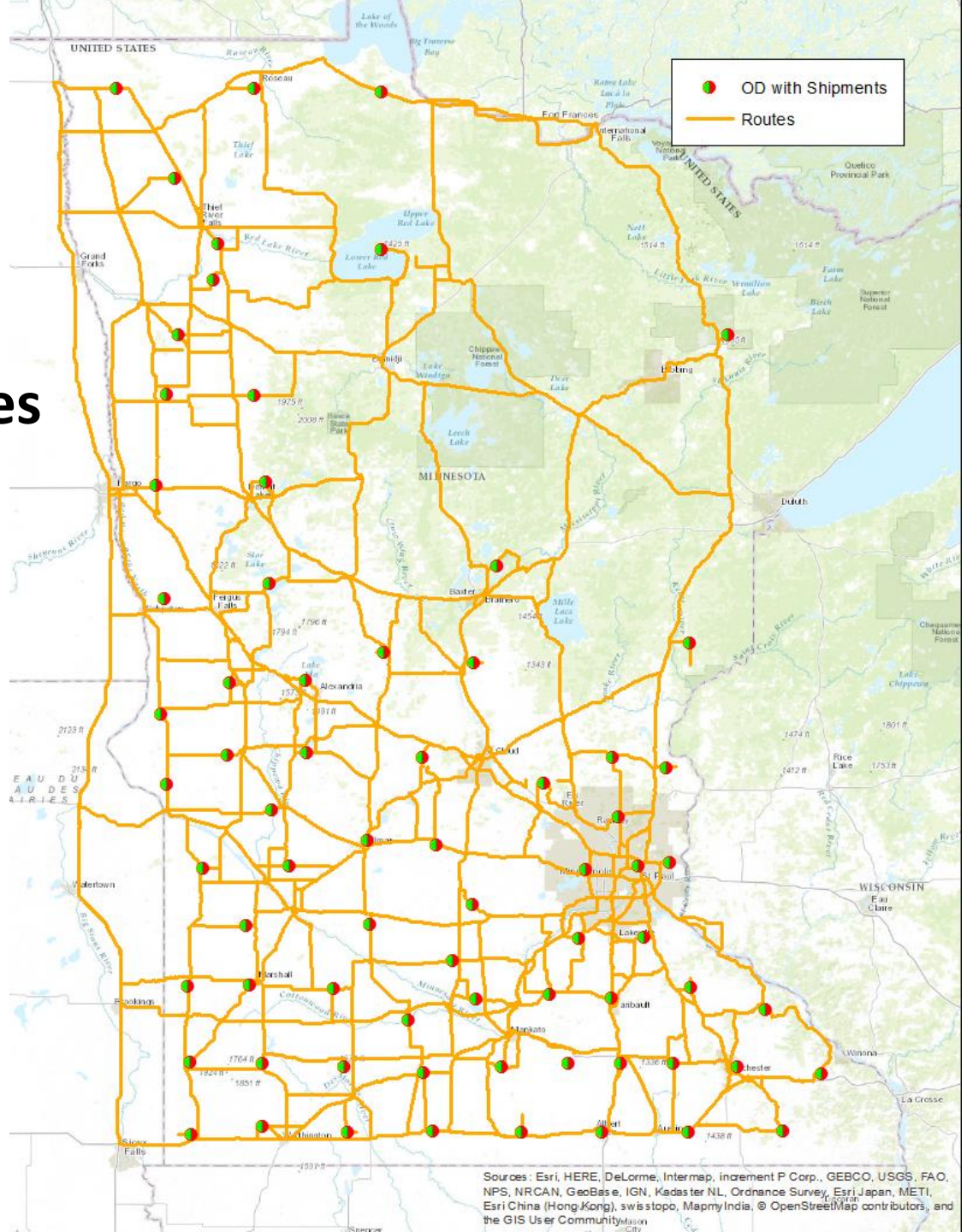
- Network Data Description
  - (87 MN counties) * (87 MN counties) = 7,569 total O-D routes
  - Appended shipment data to these O-D routes (4,489 routes)
  - Merged these O-D routes into one “flattened” dataset
  - Joined the merged O-D routes with 30,389 MN road segments
  - Resultant layer contains 30,389 road segments containing **road usage** and **shipment weight totals**



Evolving Supply Chains and Local Freight Flows:
GIS Analysis of Minnesota Cereal Grain Movement

Origins, Destinations, and O-D Routes

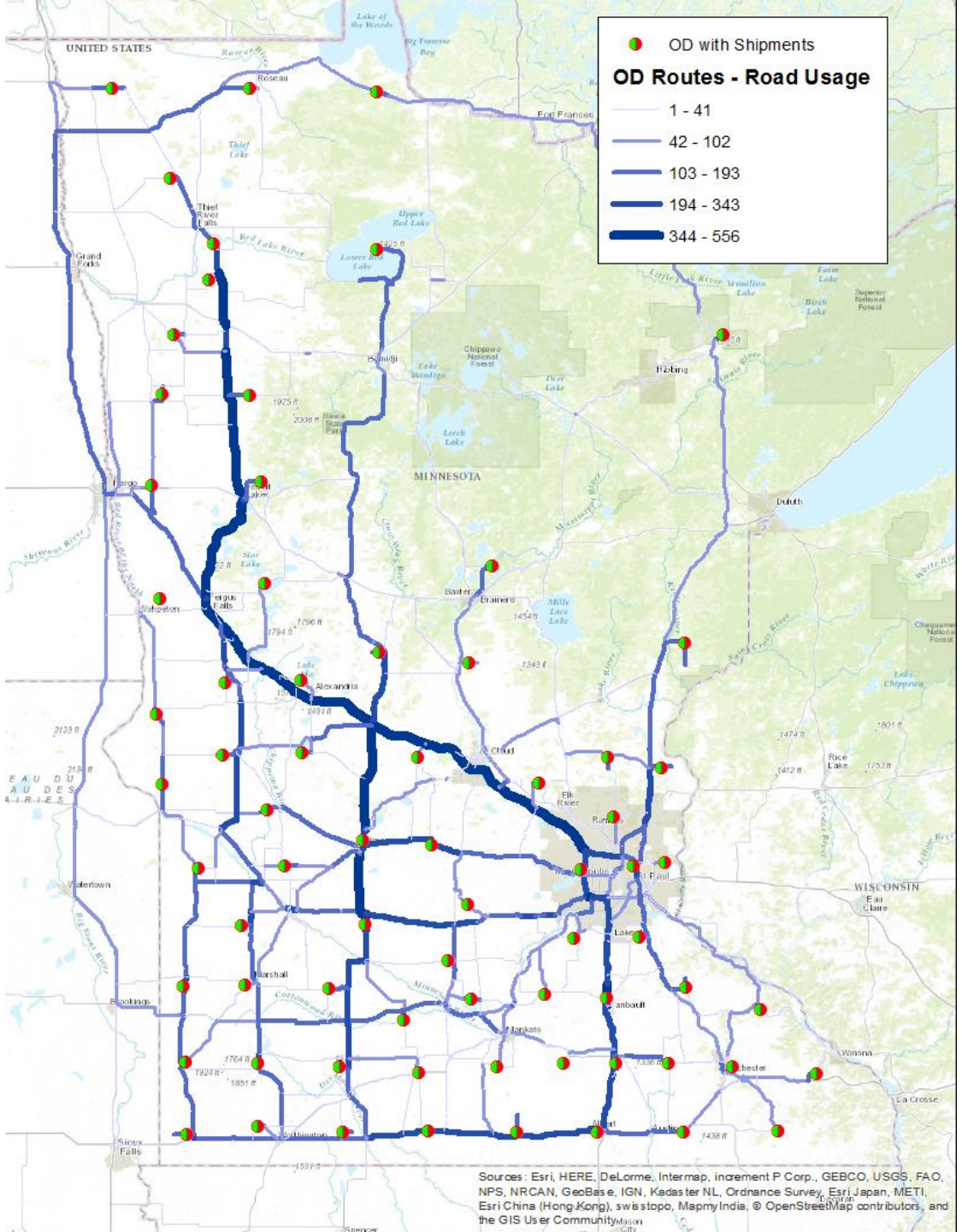
Origins (All MN Counties)
Destinations (All MN Counties)



Evolving Supply Chains and Local Freight Flows:
GIS Analysis of Minnesota Cereal Grain Movement

Road Usage

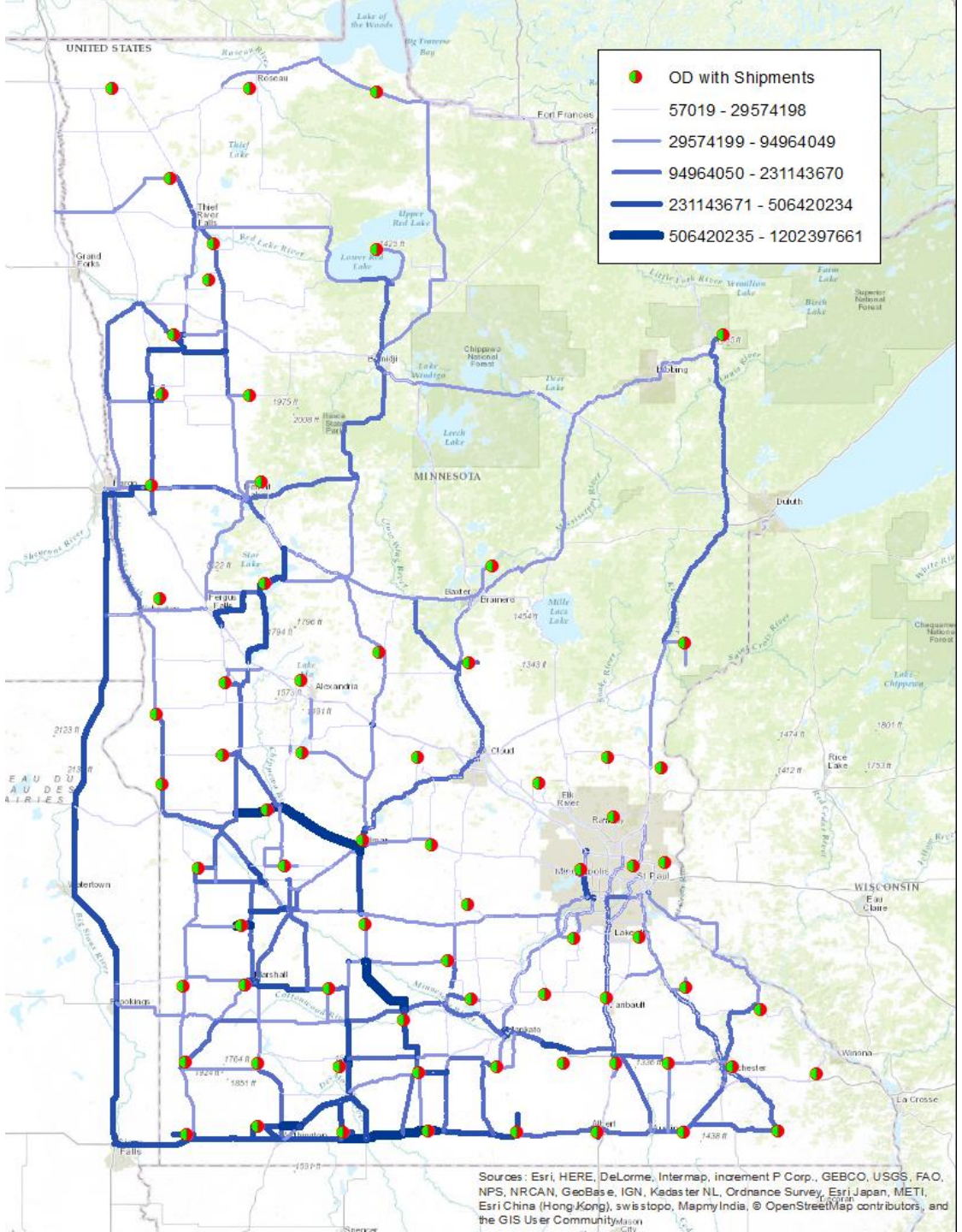
Origins (All MN Counties)
Destinations (All MN Counties)



Evolving Supply Chains and Local Freight Flows:
GIS Analysis of Minnesota Cereal Grain Movement

Shipping Weight

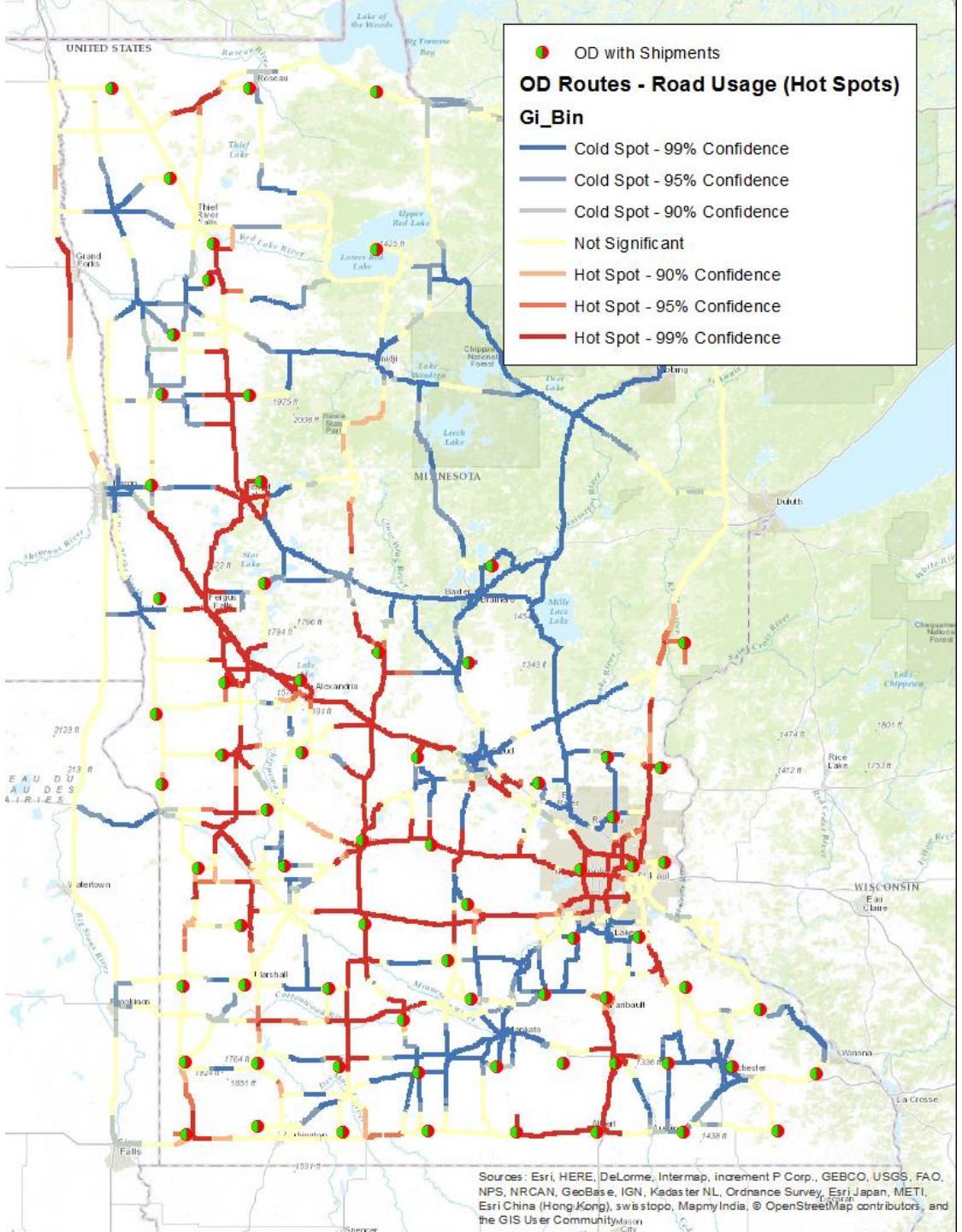
Origins (All MN Counties)
Destinations (All MN Counties)



Evolving Supply Chains and Local Freight Flows:
GIS Analysis of Minnesota Cereal Grain Movement

Road Usage (Hot Spots)

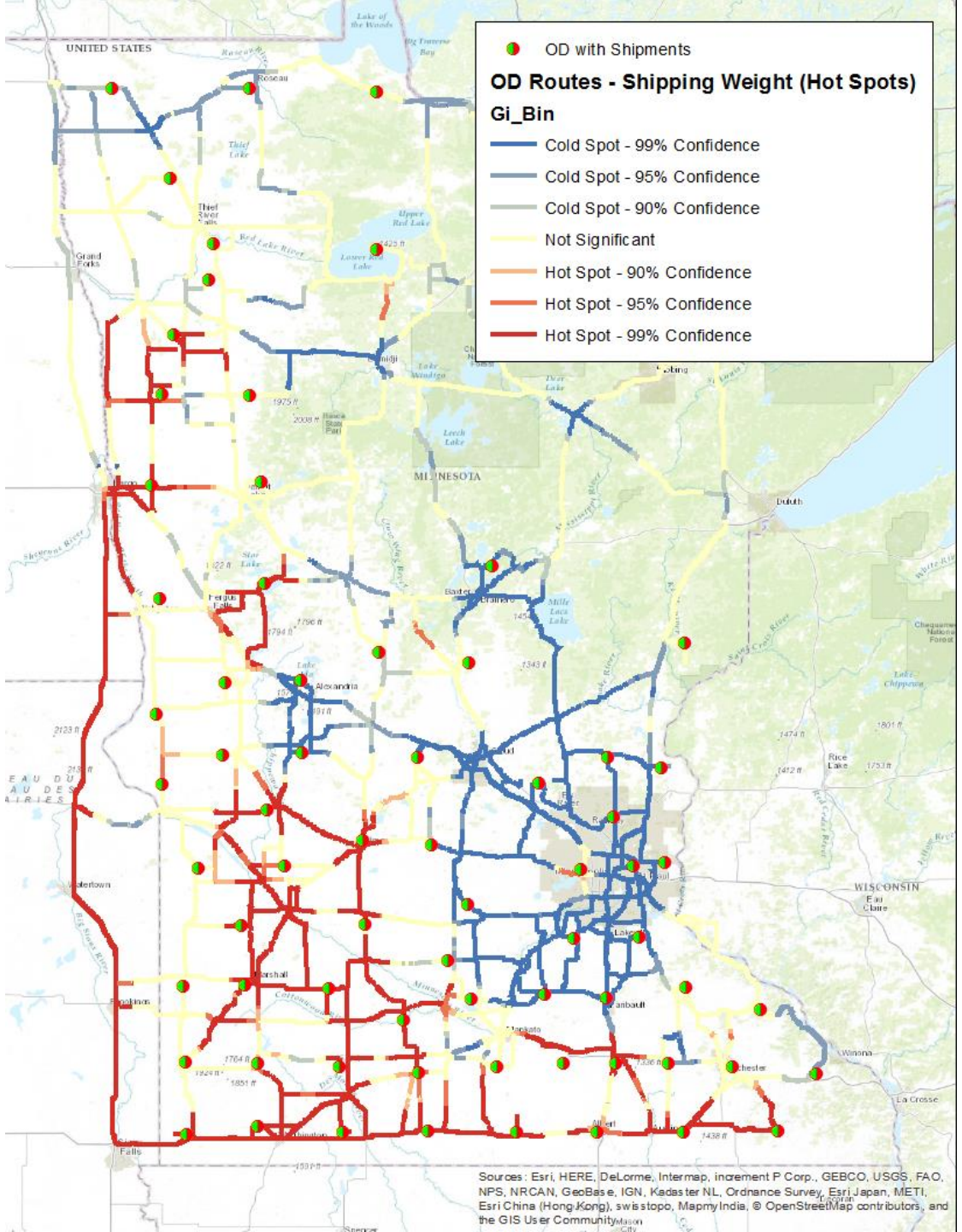
Origins (All MN Counties)
Destinations (All MN Counties)



Evolving Supply Chains and Local Freight Flows:
GIS Analysis of Minnesota Cereal Grain Movement

Shipping Weight (Hot Spots)

Origins (All MN Counties)
Destinations (All MN Counties)

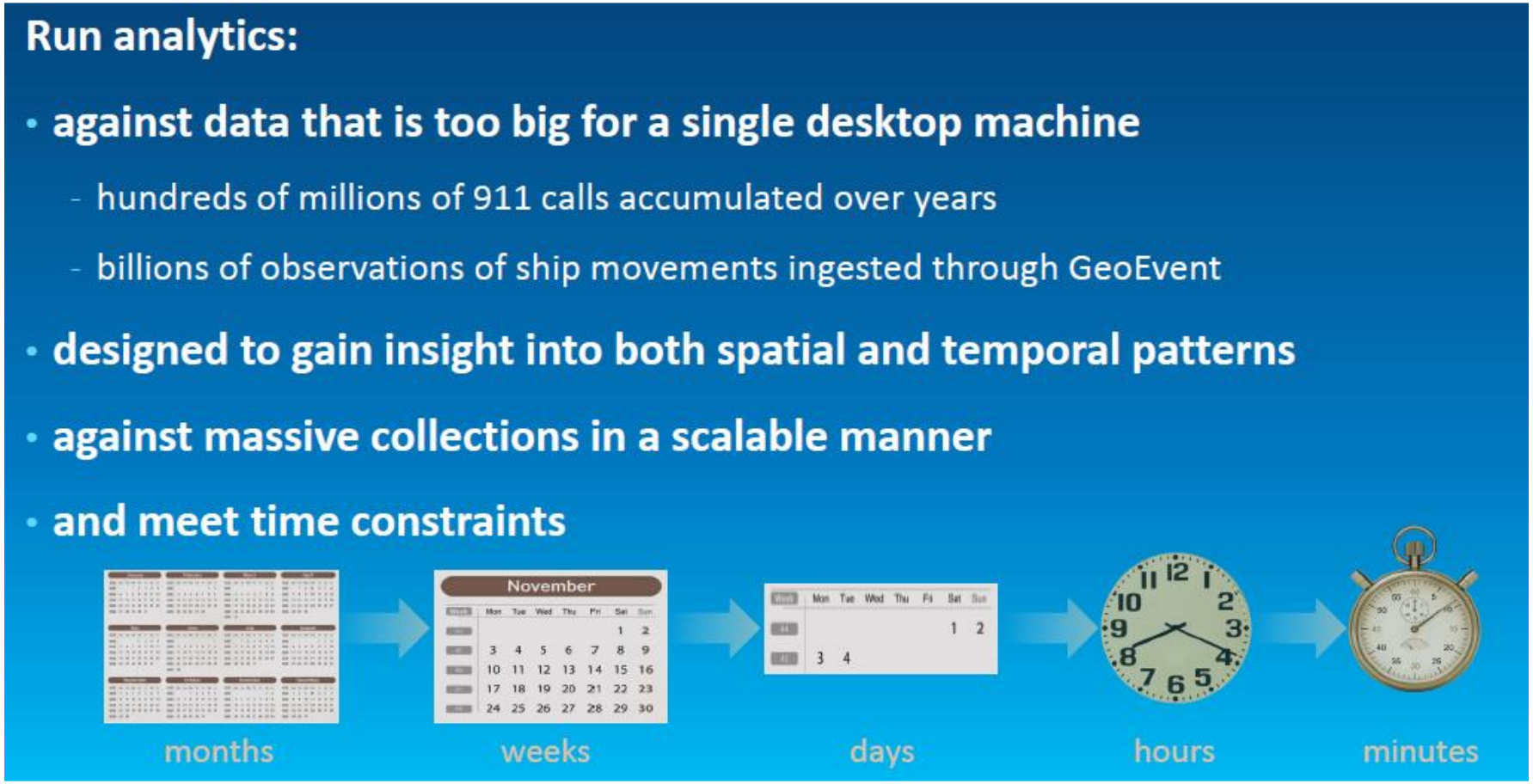


Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- Future Work / Analysis
  - Additional Industry Clusters (e.g., Agricultural Products, Animal Products, Mining Products, etc.) and Commodity Shipment Types (e.g., Processed Food Products, Dimension Stone, etc.)
  - More granular unit of analysis (1,031 MN zip codes) * (1,031 MN zip codes) = 1,062,961 total O-D routes
- Devise streamlined, and faster, data processing workflows…

Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

- Issue / Need

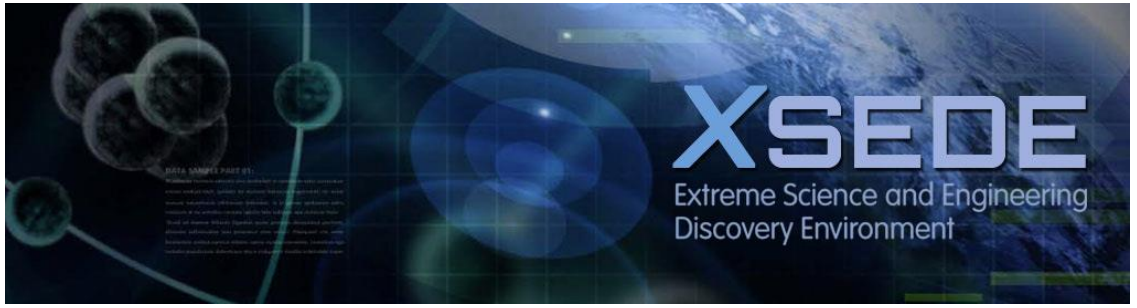


Evolving Supply Chains and Local Freight Flows: GIS Analysis of Minnesota Cereal Grain Movement

Developing novel spatial and spatiotemporal methods that can take advantage of newly emerging data-intensive computational resources.

- ArcGIS Enterprise
- XSEDE and GIS (http://www.gisandbox.org)

SDSC SAN DIEGO SUPERCOMPUTER CENTER



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Domain – Spatial Justice / Spatial Equality

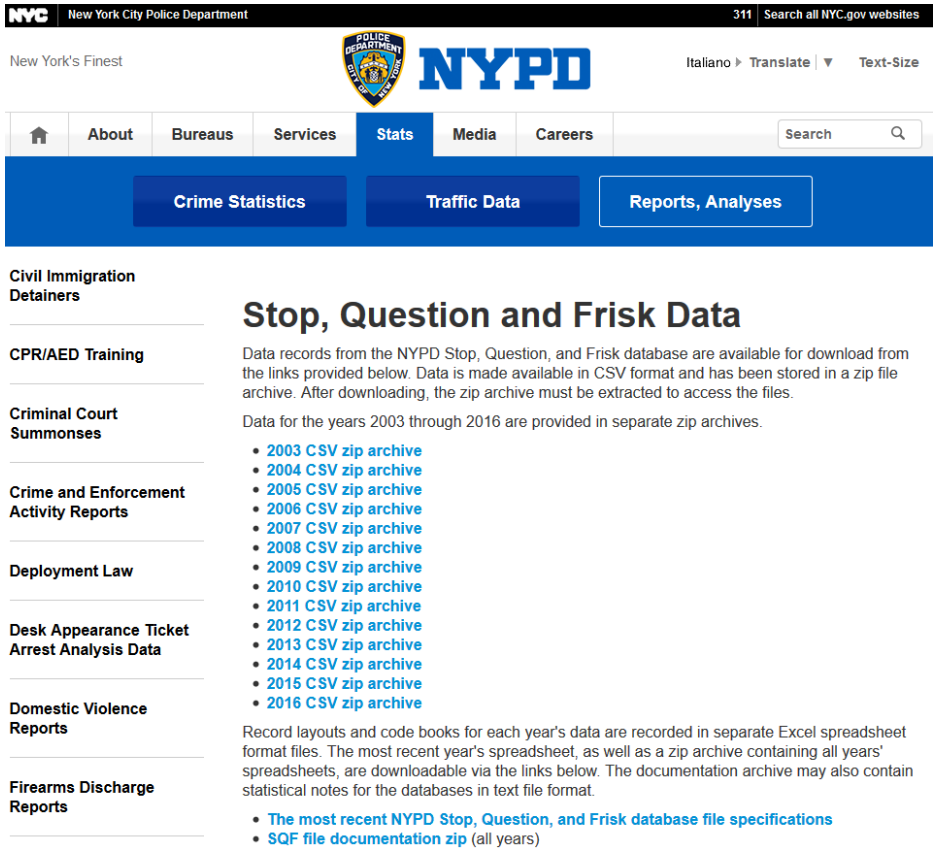
- Research Question – “Does the race or ethnicity of an individual being stopped by a police officer have a significant role in an individual being frisked and by how much?”

Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Policy encourages police officers to stop people they deem suspicious, question them, and to frisk them for drugs, contraband, or weapons if illegal activities are suspected.
  - Reasonable suspicion is the belief that someone poses a dangers, has committed a crime, or is about to commit a crime.
- Race cannot be a factor for the frisk.
- The New York City Stop-and-Frisk Policy is an example of how a policy intended to keep the public safe, now has a negative public perspective.

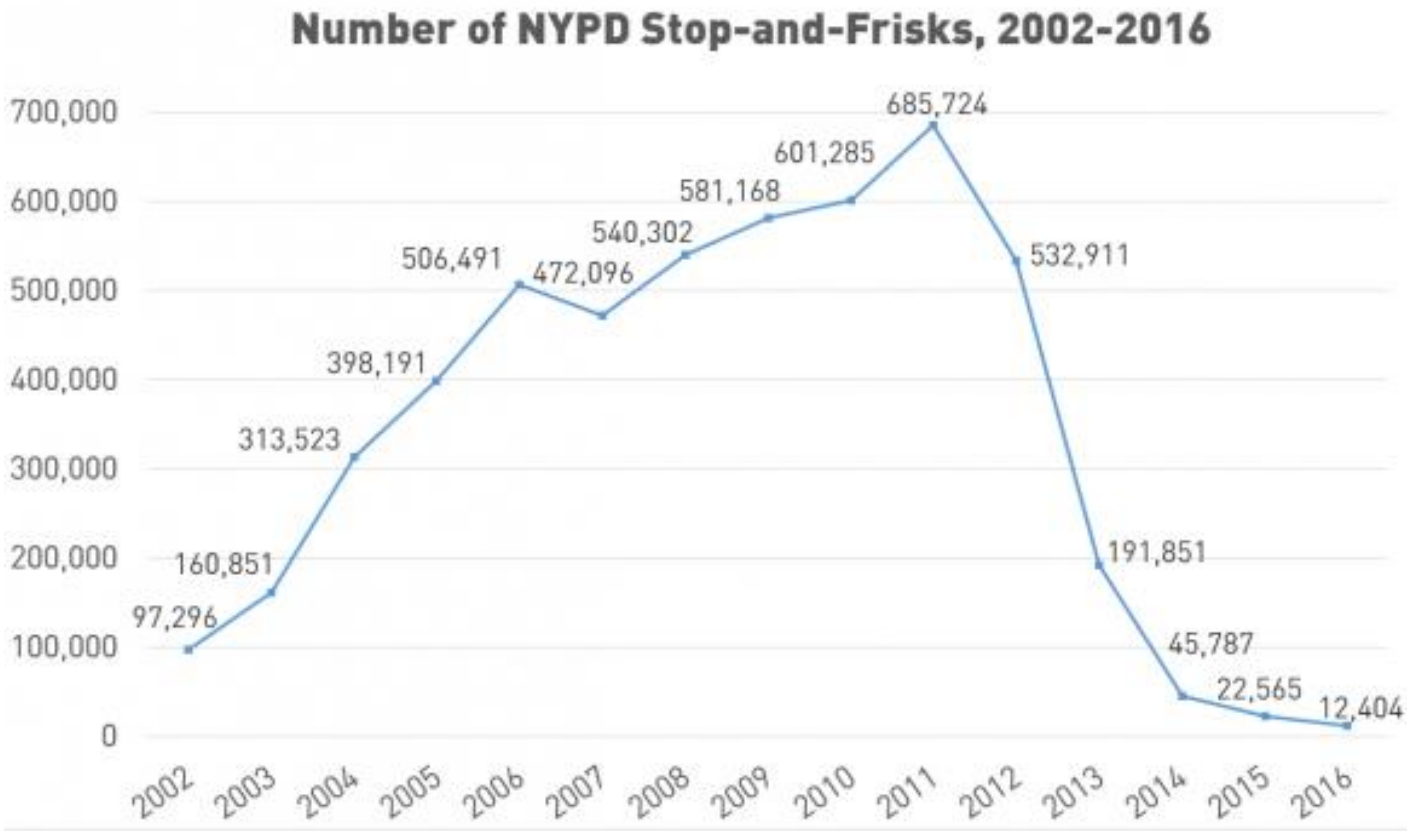
Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Data Provider
  - Stop-and-Frisk data records are available from the NYPD Stop, Question, and Frisk database.



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

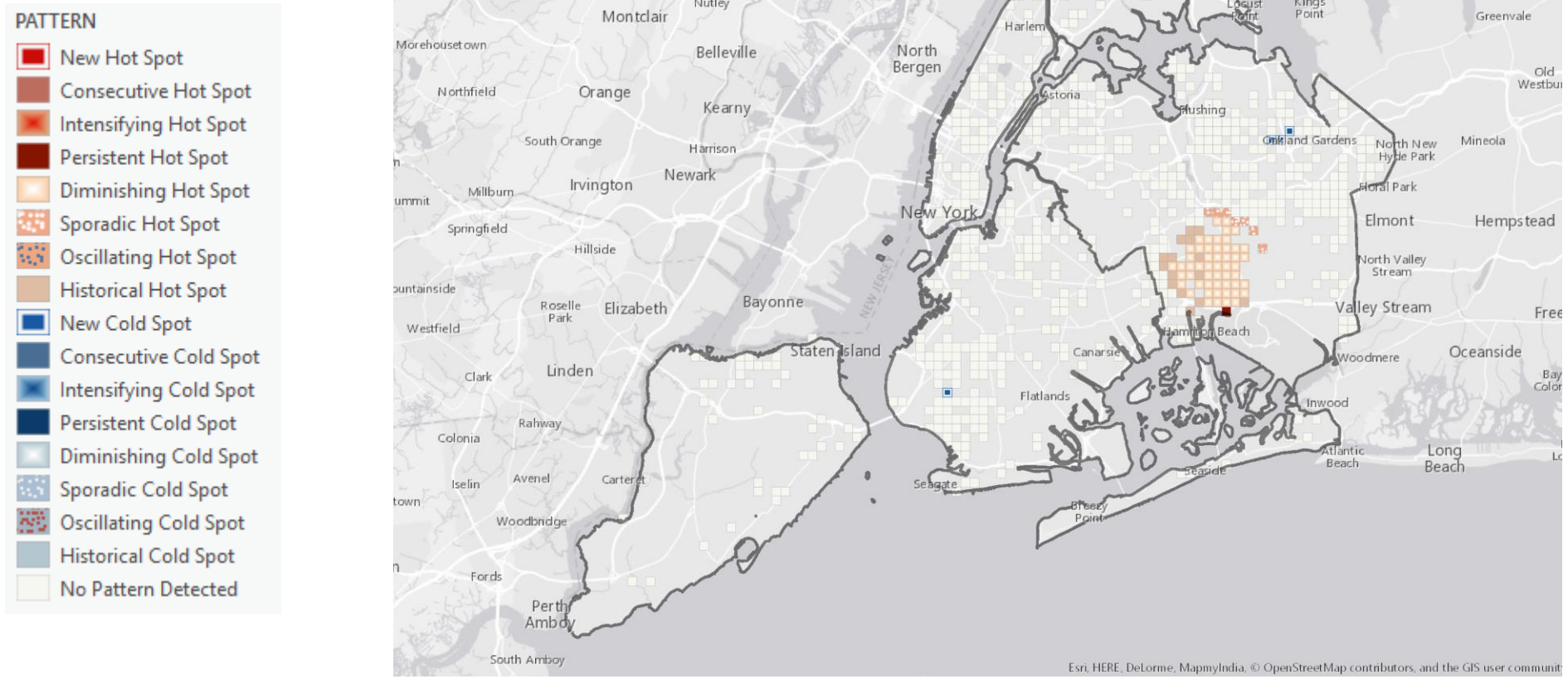
- Stop-and-Frisk Data Description
  - 5,162,445 New York City police stops and street interrogations (2002-2016)



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Asian
- Stop-and-Frisk
- 2014

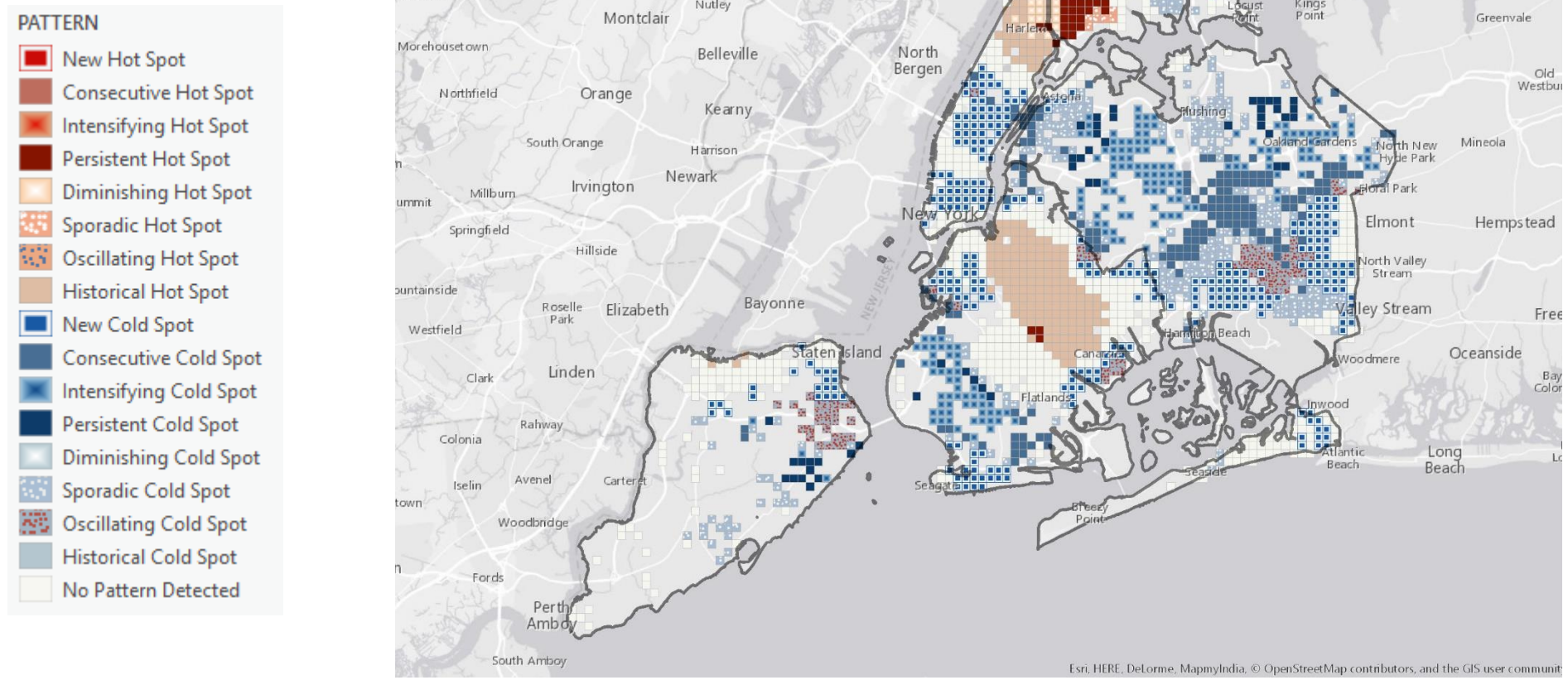
Emerging Hot Spot Analysis
Spatiotemporal Trends



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Black
- Stop-and-Frisk
- 2014

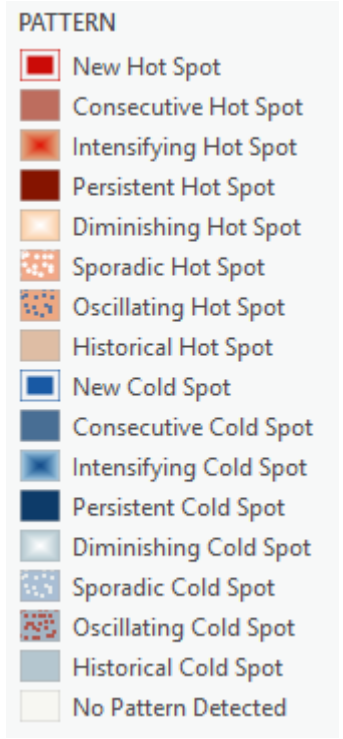
Emerging Hot Spot Analysis
Spatiotemporal Trends

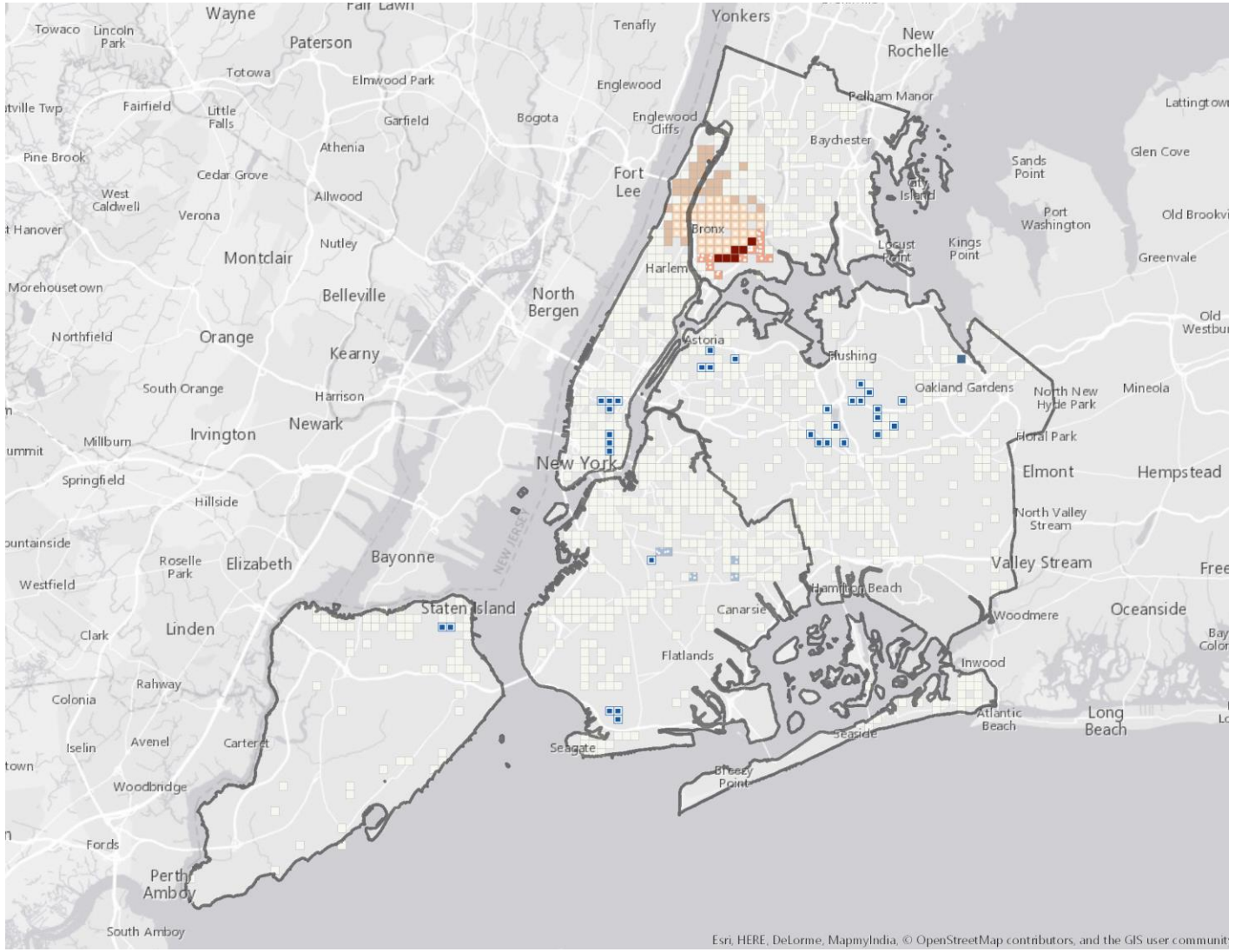


Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Black Hispanic
- Stop-and-Frisk
- 2014

Emerging Hot Spot Analysis
Spatiotemporal Trends

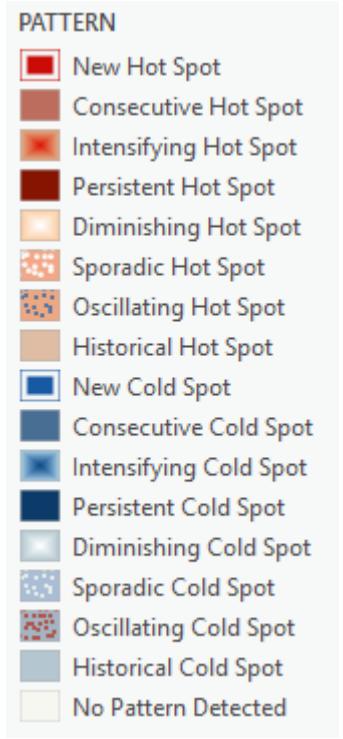


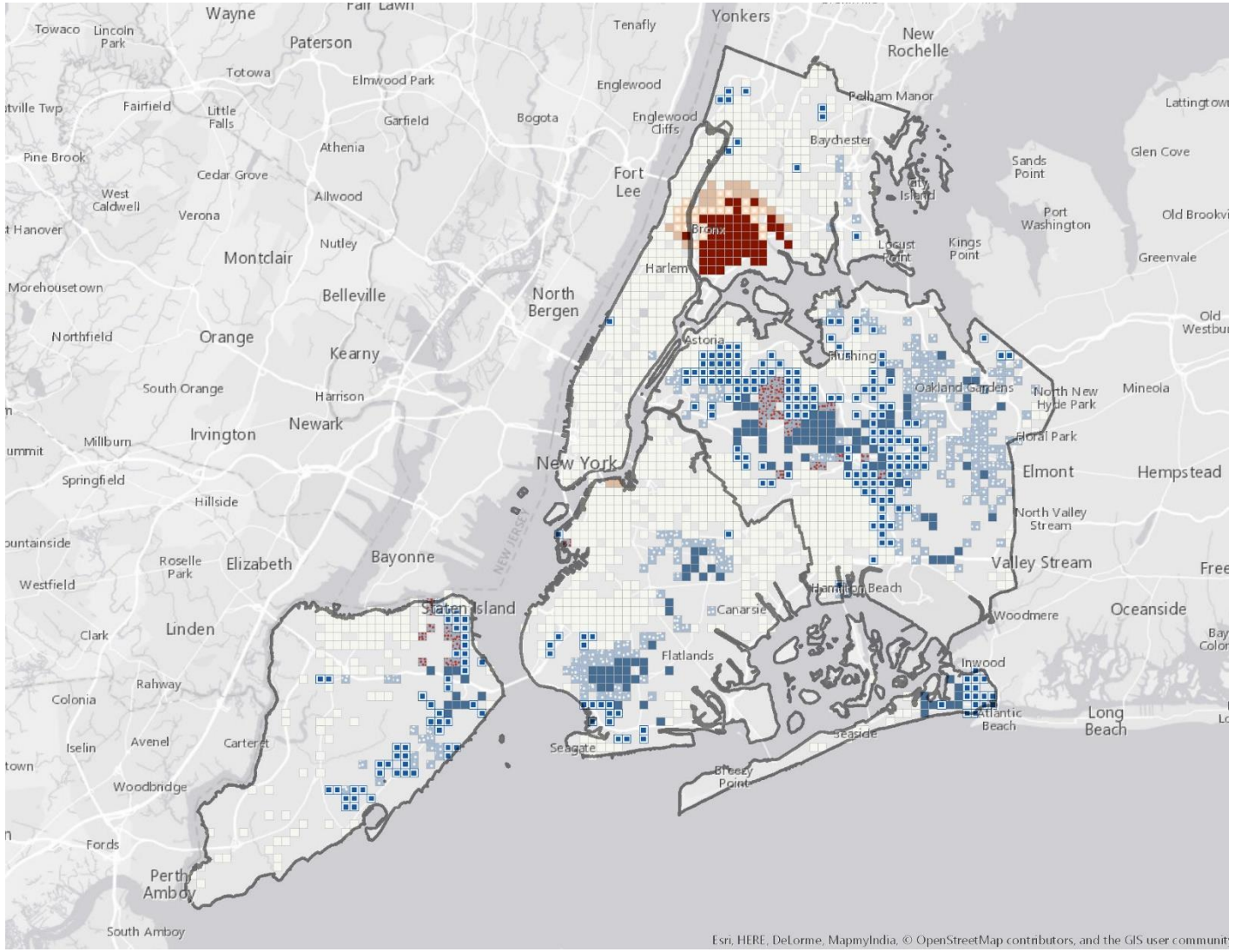


Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- White Hispanic
- Stop-and-Frisk
- 2014

Emerging Hot Spot Analysis
Spatiotemporal Trends

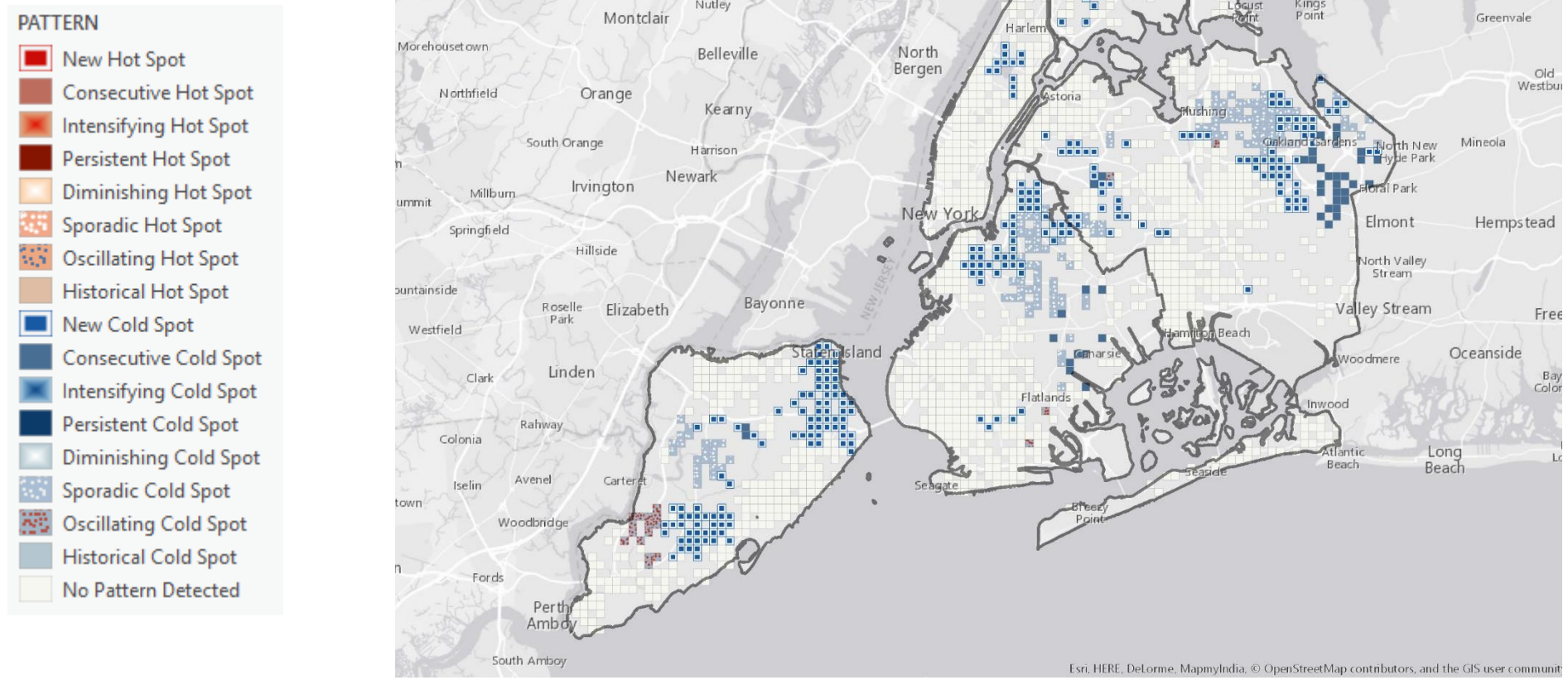




Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

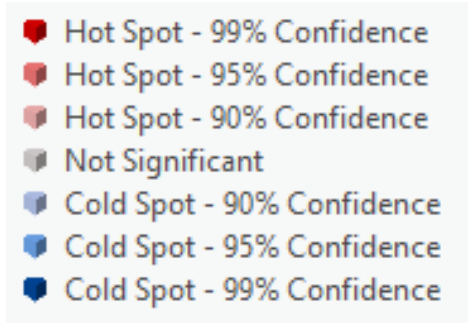
- White
- Stop-and-Frisk
- 2014

Emerging Hot Spot Analysis
Spatiotemporal Trends

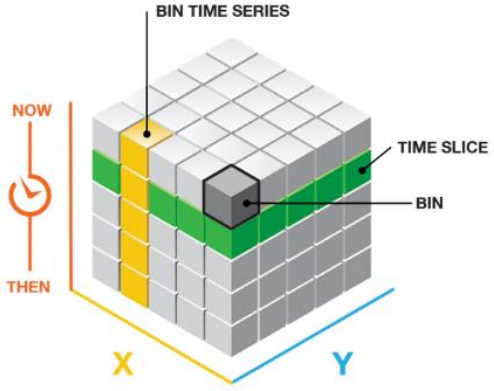


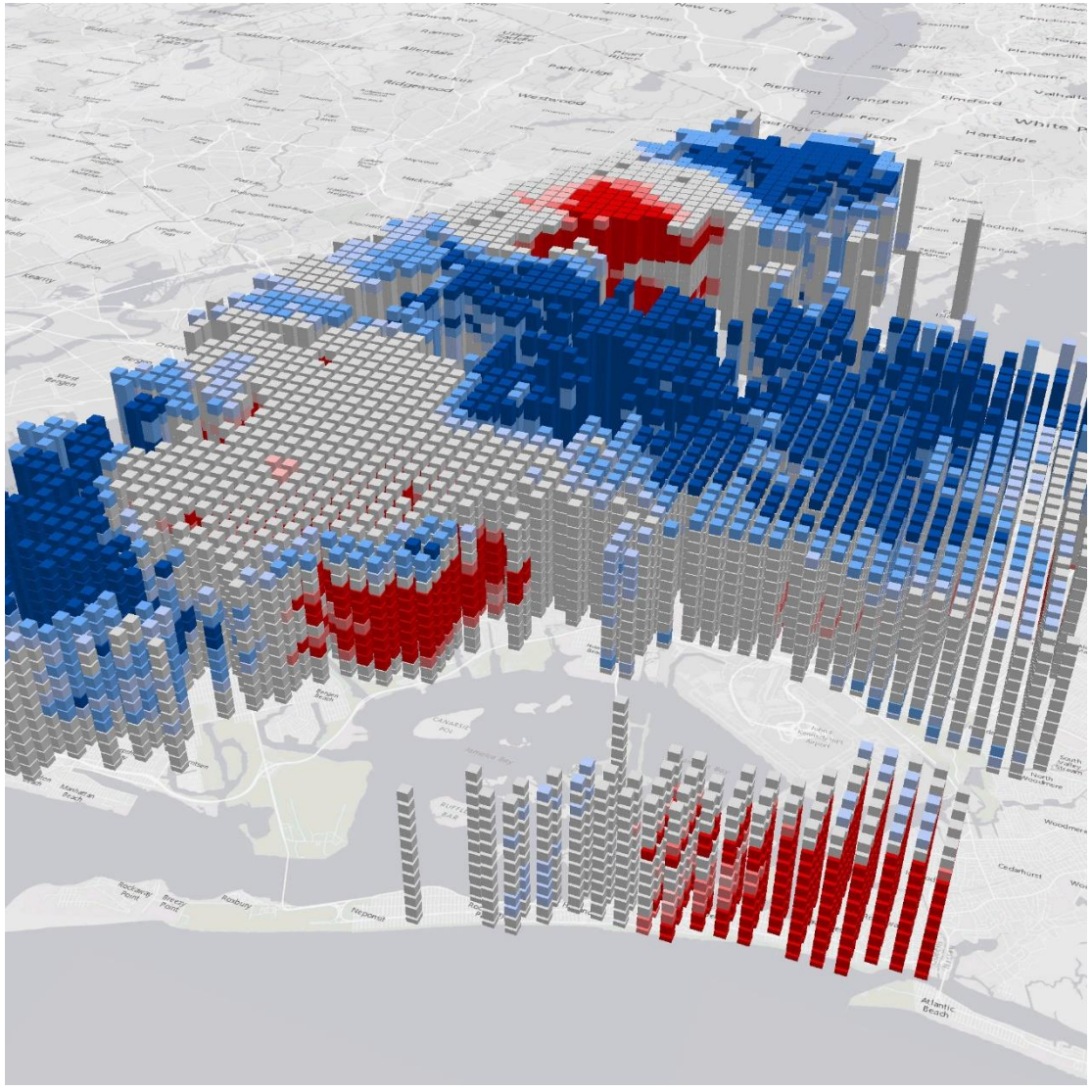
Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Black
- Stop-and-Frisk
- 2014



Emerging Hot Spot Analysis
Visualize Space-Time Cube



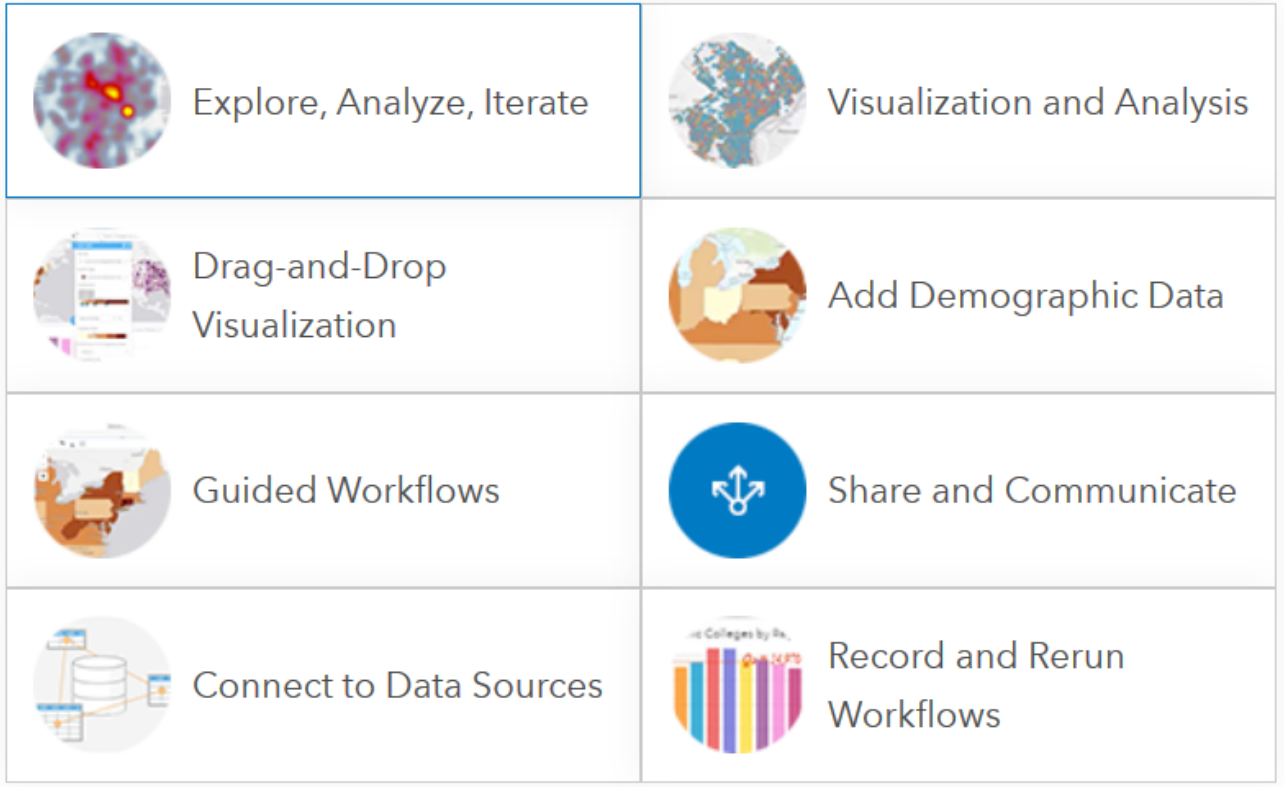


Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

- Future Work / Analysis
  - Conduct multiple, spatiotemporal analyses (e.g., across years, specific months, days, hours, specific attribute types, and combinations of these) for NYC.
  - Examine this issue in other cities (e.g., Philadelphia, Chicago, Los Angeles)
- Devise streamlined, shareable, analytic processing workflows…

Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

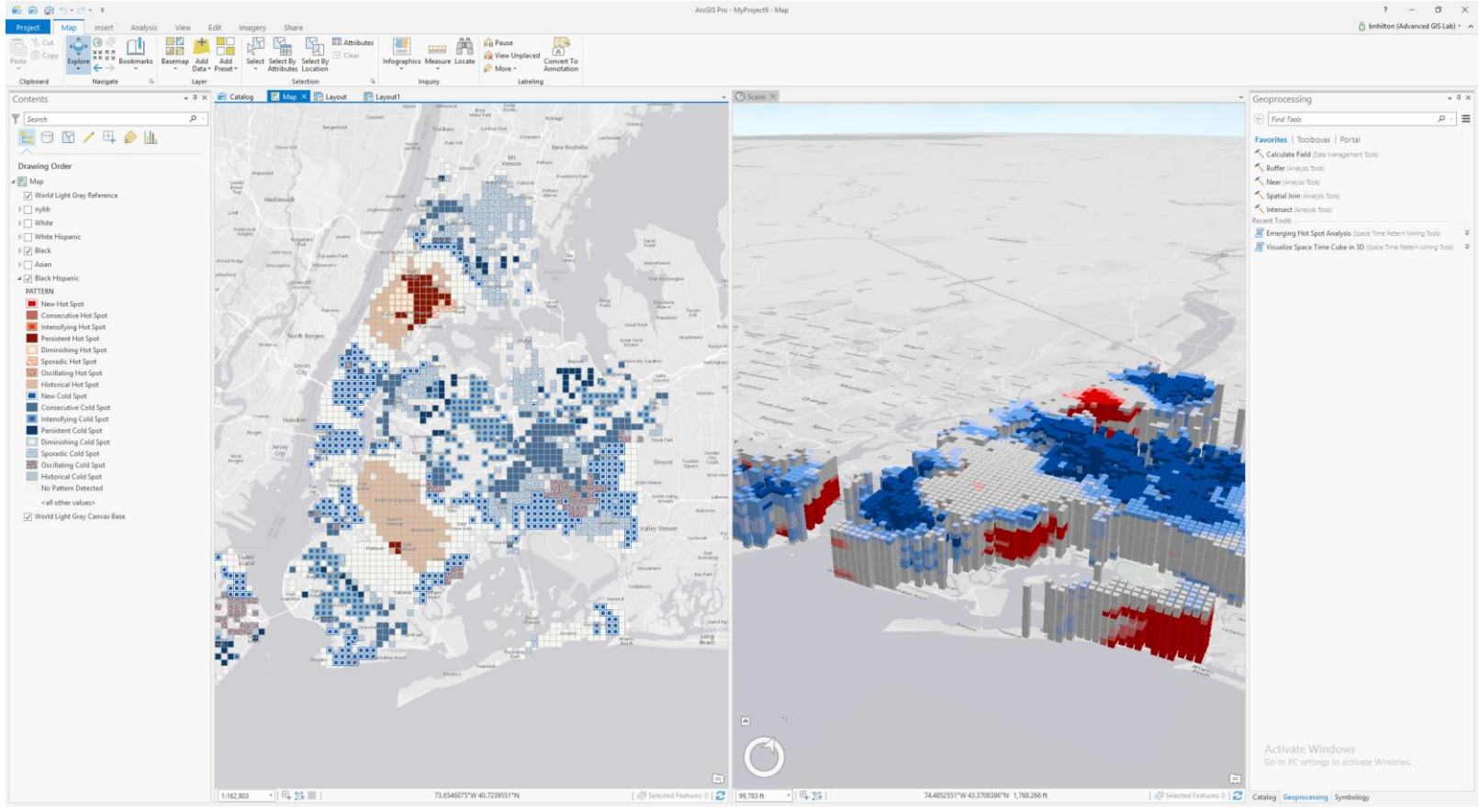
- Issue / Need
  - Web-based, data analytics and exploration tool for conducting multiple, spatiotemporal analyses, of spatial and non-spatial data, where you can:



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

Developing methodologies to derive insight into spatial big data for inferential understanding and decision making.

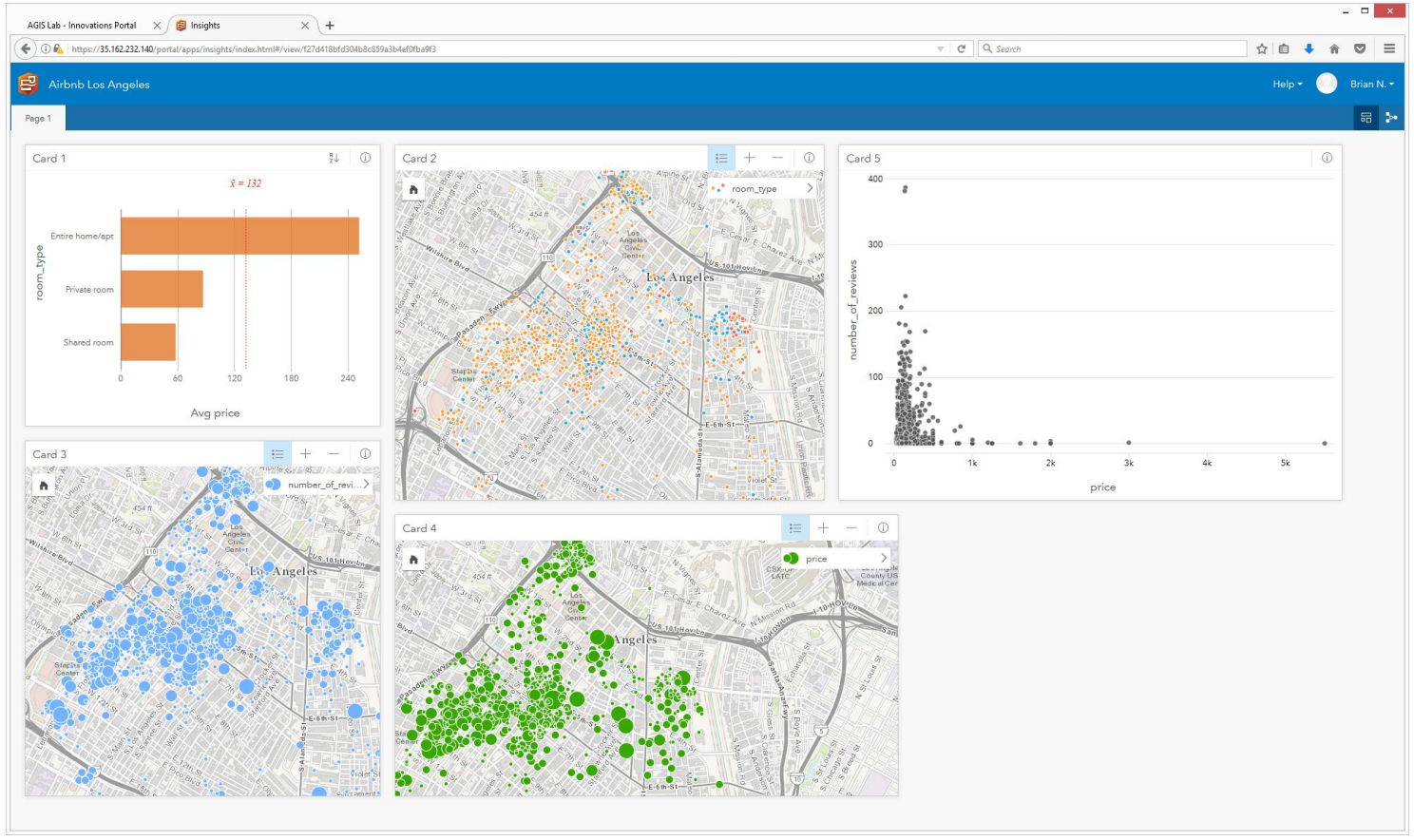
- ArcGIS Pro
  - Emerging Hot Spot Analysis
  - Space-Time Cube



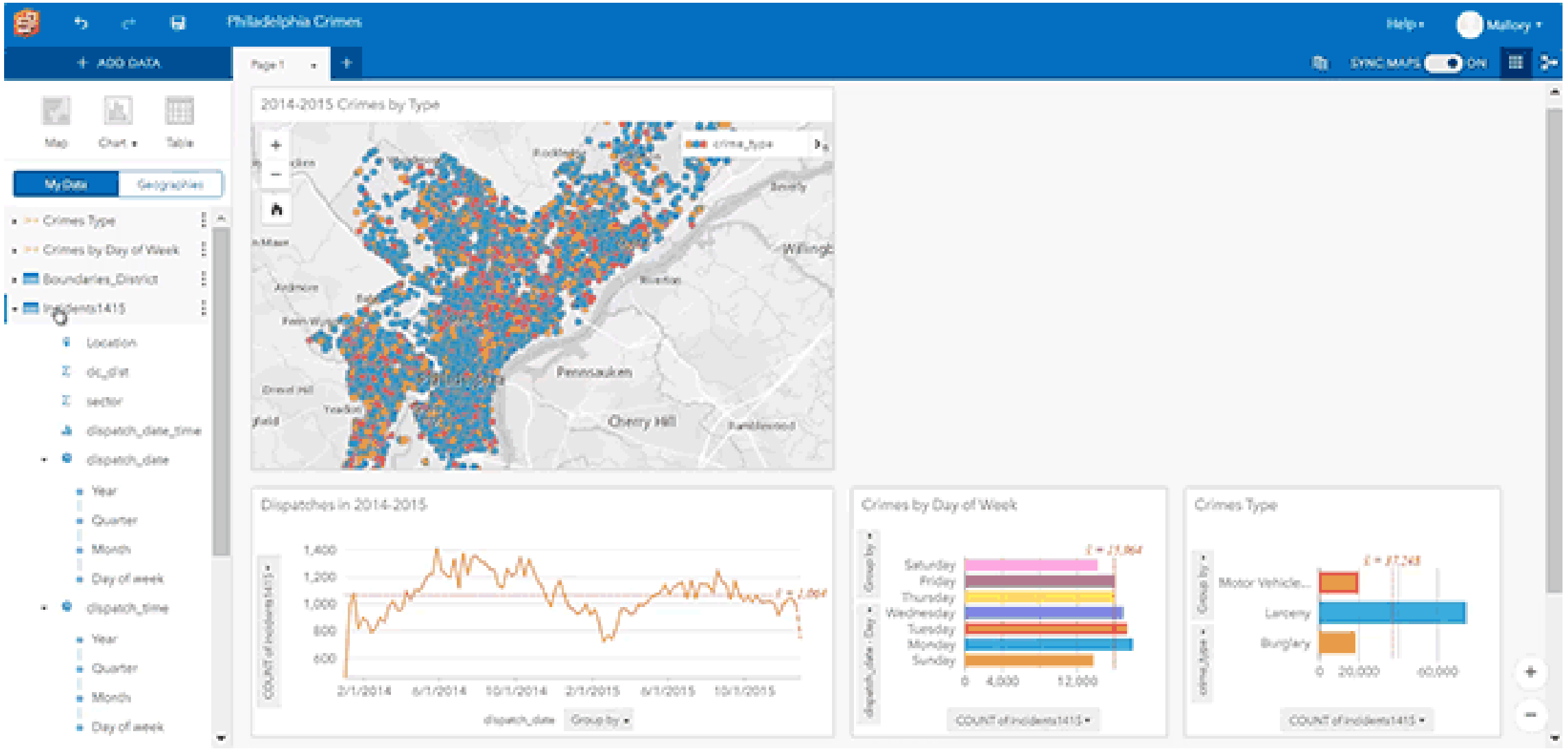
Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

Developing methodologies to derive insight into spatial big data for inferential understanding and decision making.

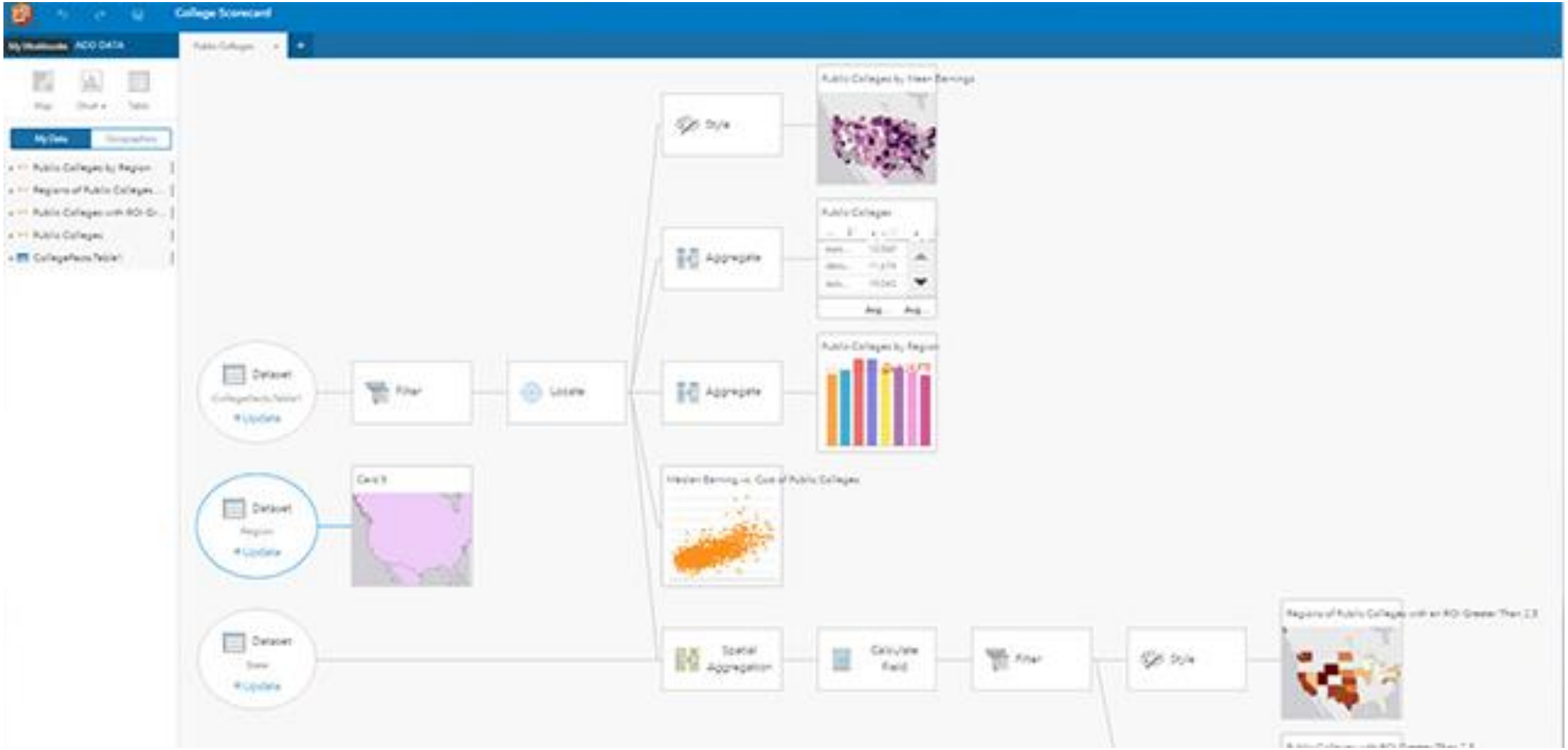
- Insights for ArcGIS
- Web-based
- Data analytics tool



Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

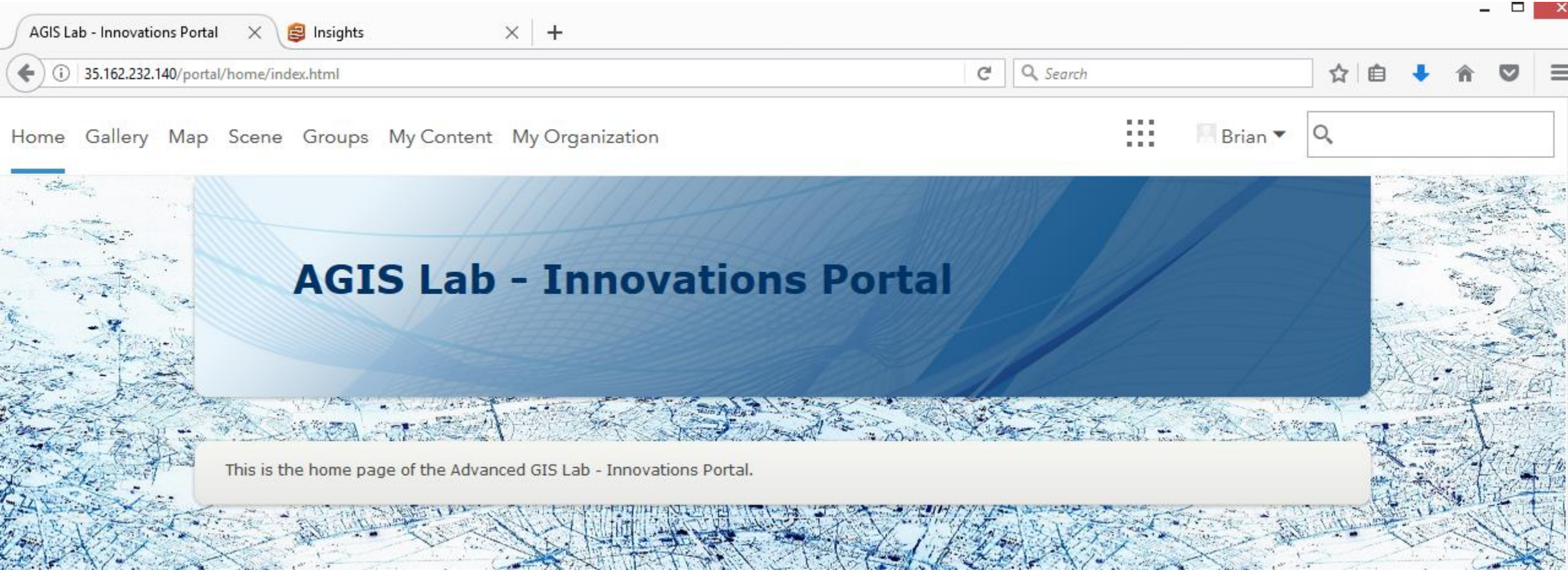


Stop-and-Frisk Policy from a Quantitative and Spatial Perspective

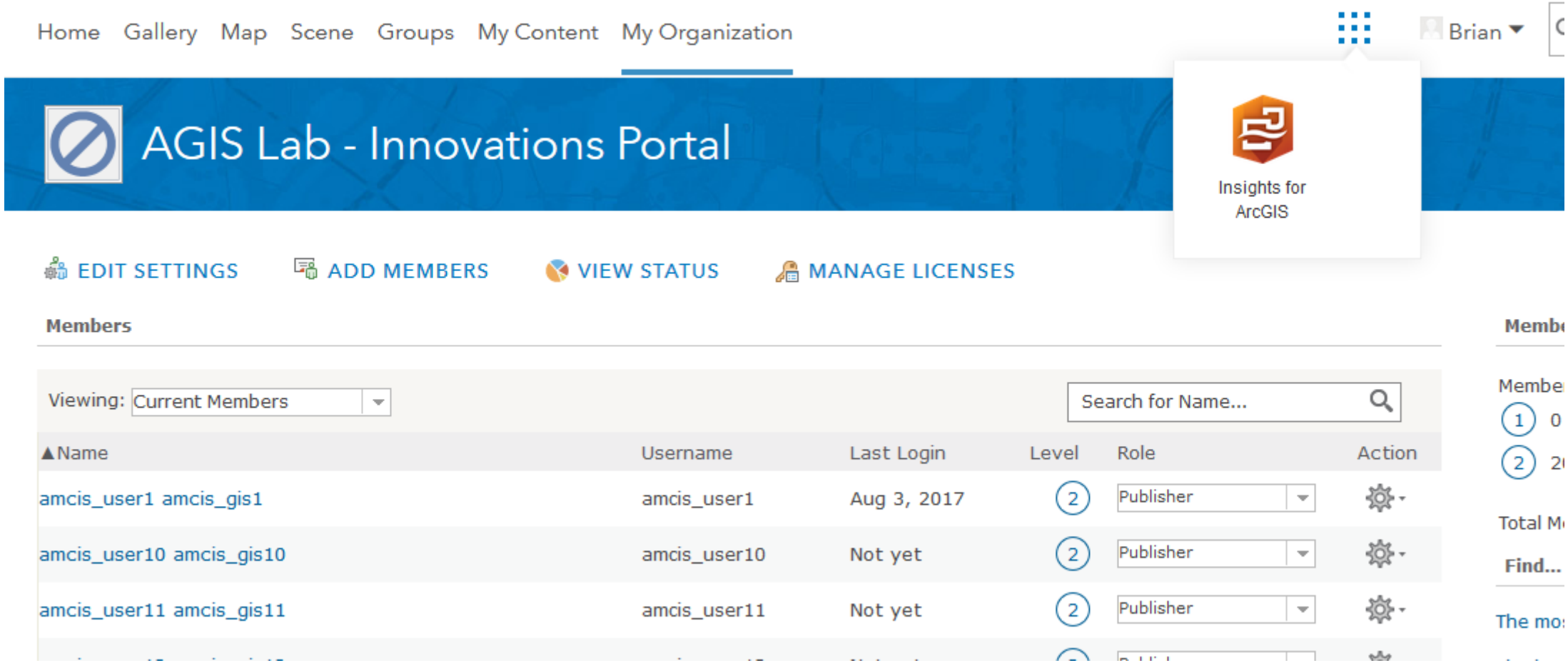


Insights for ArcGIS - Hands-on demo

http://agislab.org



Insights for ArcGIS - Hands-on demo



Appendix

Airbnb Boston

Help | Brian N.

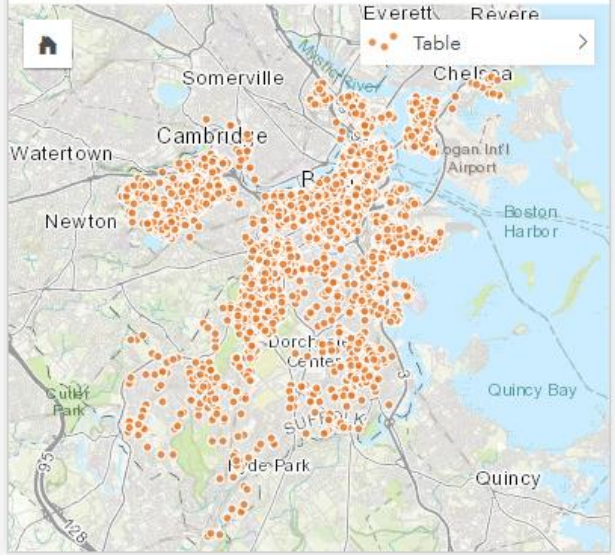
+ ADD DATA | Page 1 | +

Map | Chart | Table

Card 2

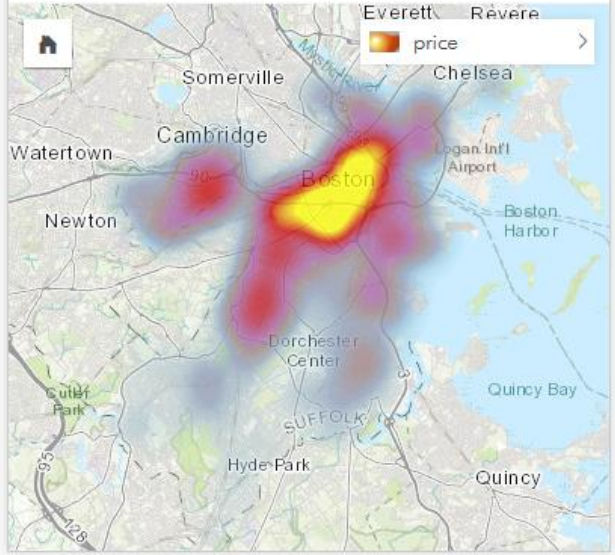
Table

Card 1



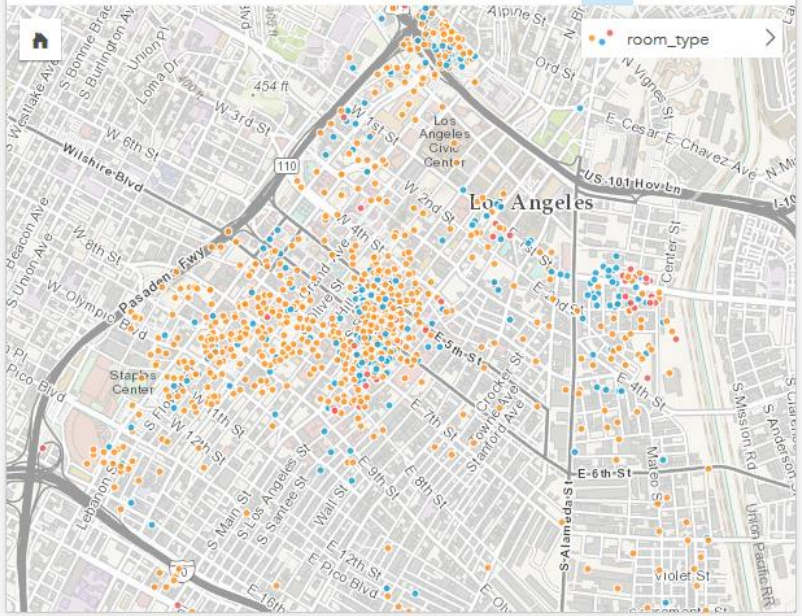
Card 2

Card 3

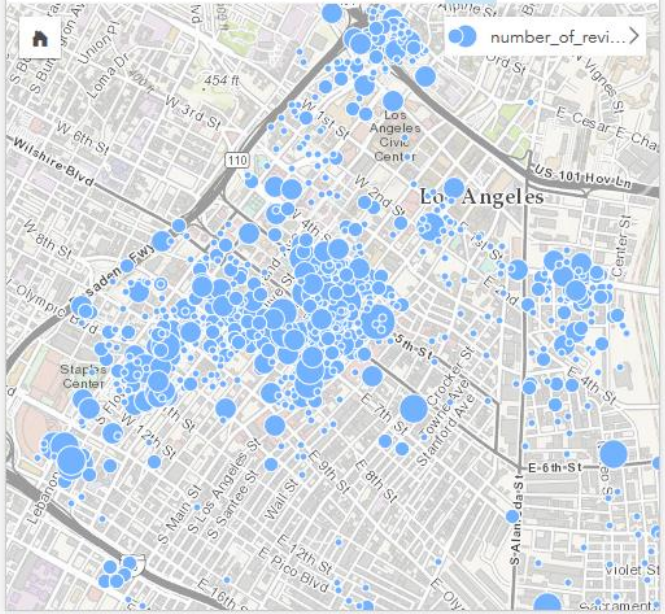


Card 1

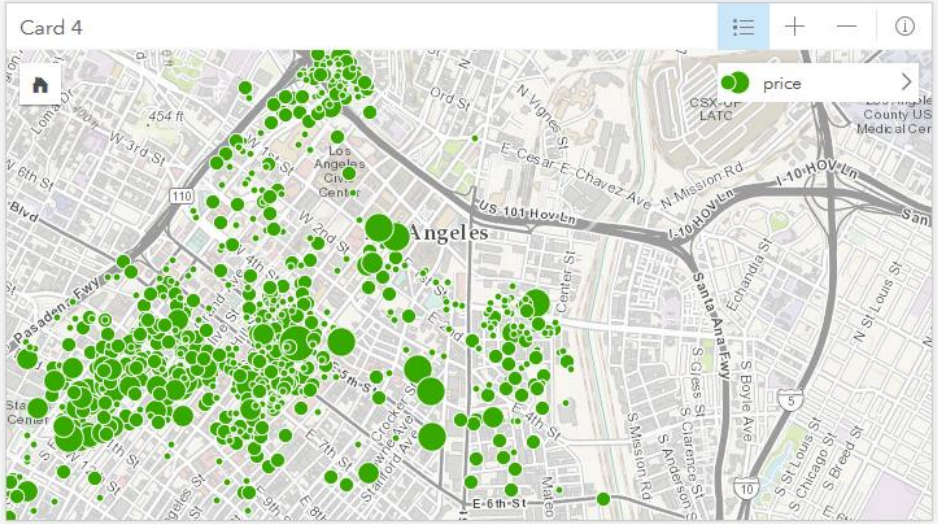
Card 2



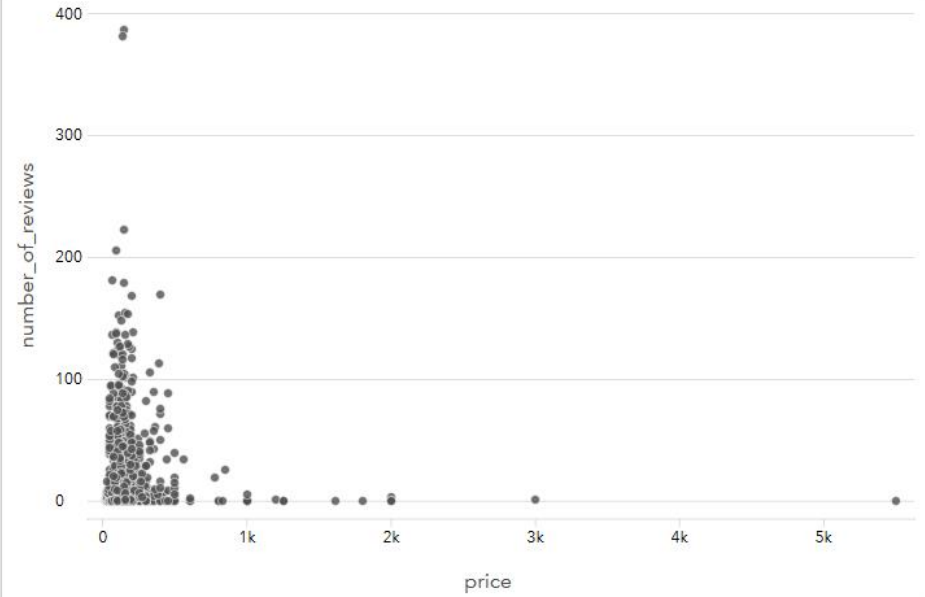
Card 3



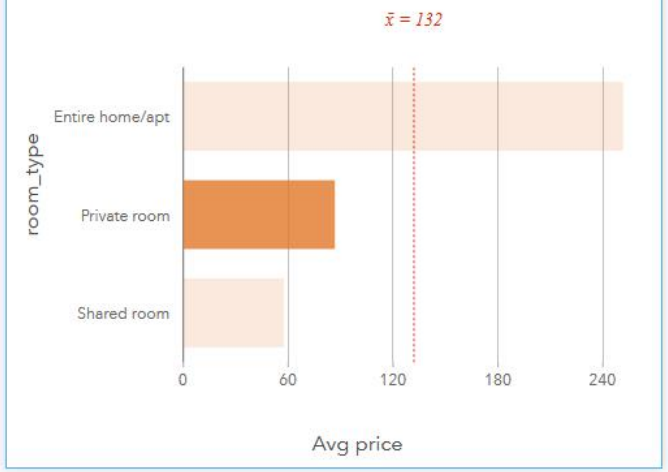
Card 4



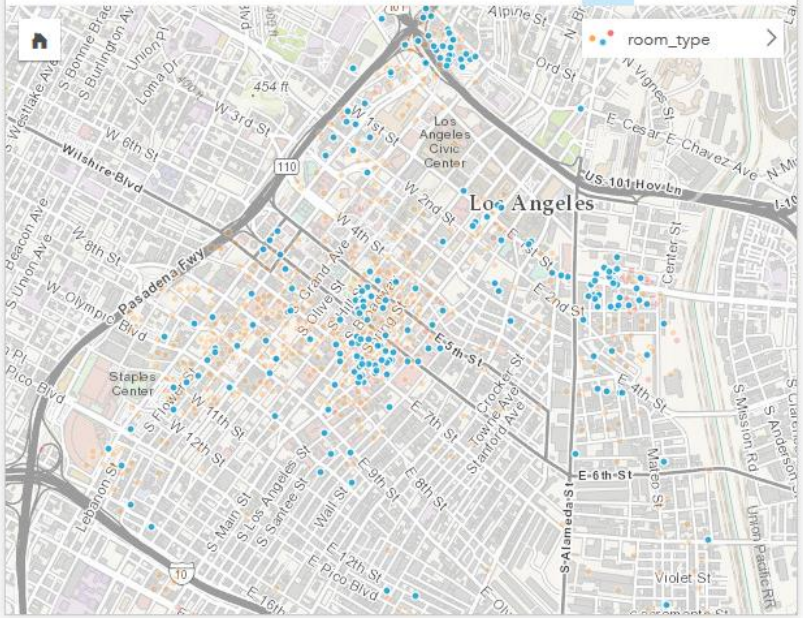
Card 5



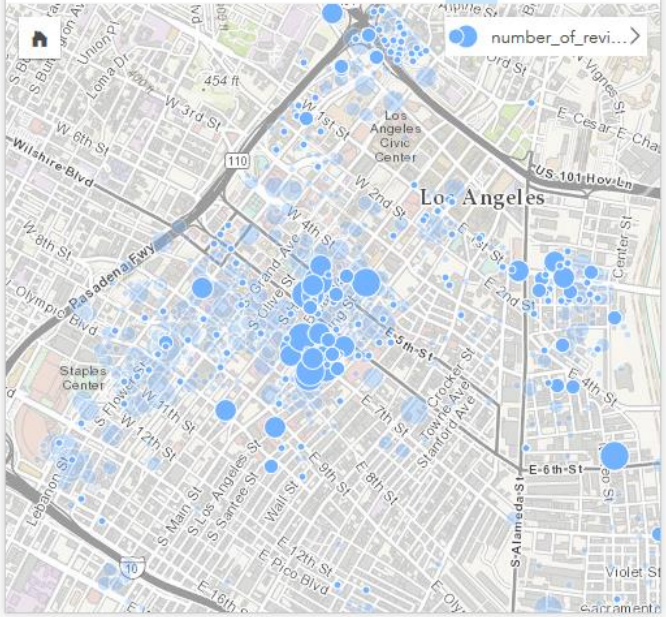
Card 1



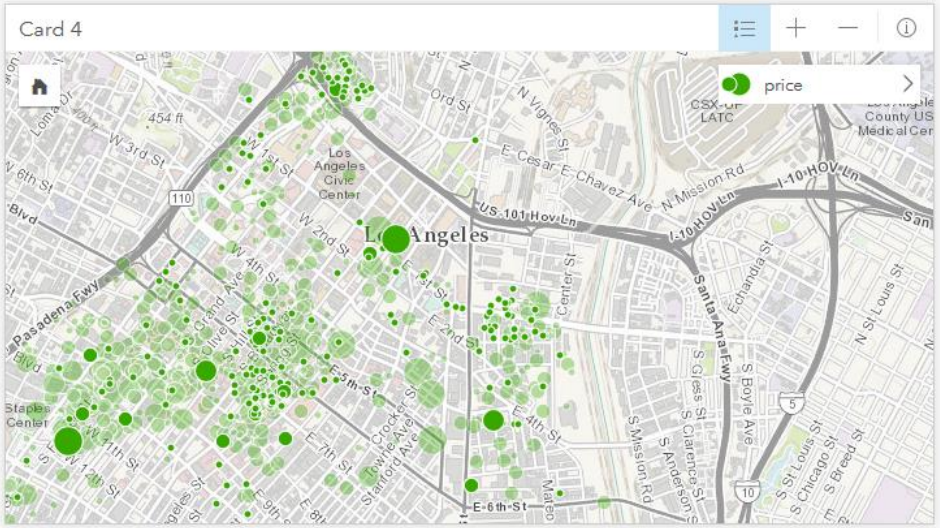
Card 2



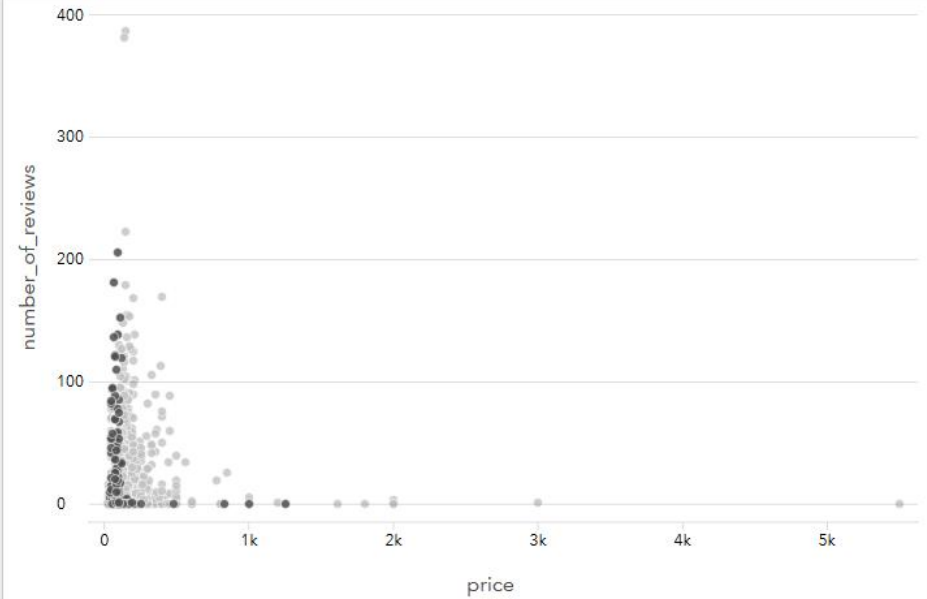
Card 3



Card 4



Card 5



Additional Resources

Developing teaching resources to better understand the use of data manipulation techniques, spatial statistics, and spatial data-mining tasks related to spatial big data.

- Jupyter Notebooks
  - https://developers.arcgis.com/python/
  - https://developers.arcgis.com/python/sample-notebooks/
  - https://notebooks.esri.com/user/VFVGul1U7slf/notebooks/samples/04_gis_analysts_data_scientists/analyze_new_york_city_taxi_data.ipynb
- MapD: https://www.mapd.com/