



GIS and Graduate IS Research

Thomas A. Horan, Ph.D.
Professor and Director
Center for Information Systems and Technology
Claremont Graduate University

GIS Workshop, ICIS Conference, Auckland New Zealand, December 14, 2014

Agenda



CGU and GIS

A Transdisciplinary Science and Practice

Overview of GIS/IS Concentration

Current Research Streams

Future Directions

Historical Background of CGU



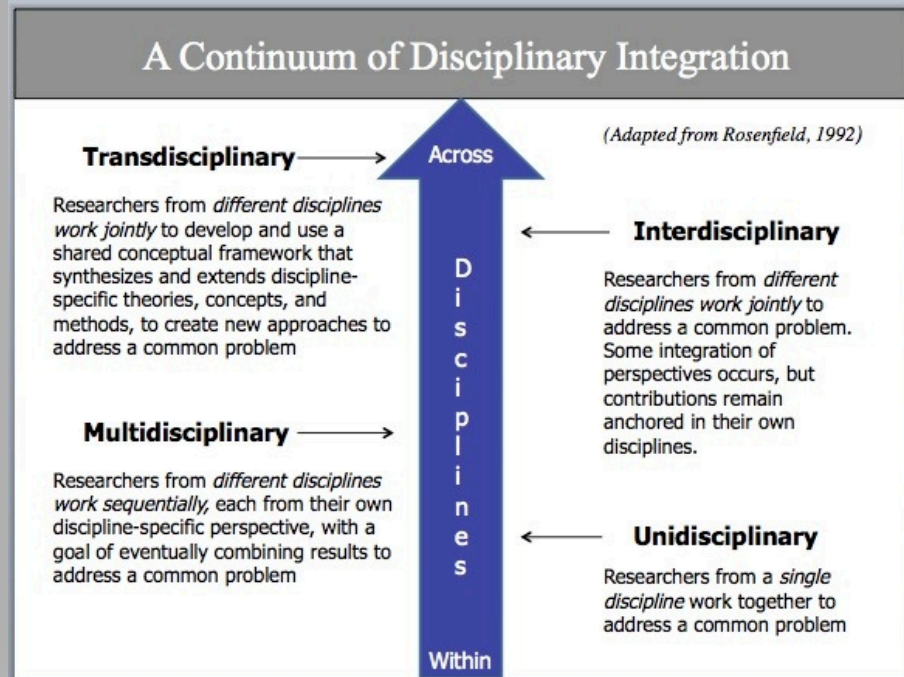
- Founded in 1925. The first university in North America to devote itself entirely to graduate study.
- Includes four Schools, and two Centers. Has established an intimate, student-focused environment designed to promote creative collaboration with faculty and fellow students.
- Part of the Claremont Colleges Consortium

About CISAT



- The Center for Information Systems and Technology (CISAT) was founded in 1983 by Dr. Paul Gray, an influential pioneer and world-renowned scholar in the field of information systems and technology (IS&T).
- CISAT takes an applied transdisciplinary approach to IS& T, with multiple concentration options in both master's and doctoral program, as well as interfield options with any of the CGU schools.

GIS as a Transdisciplinary Science and Practice



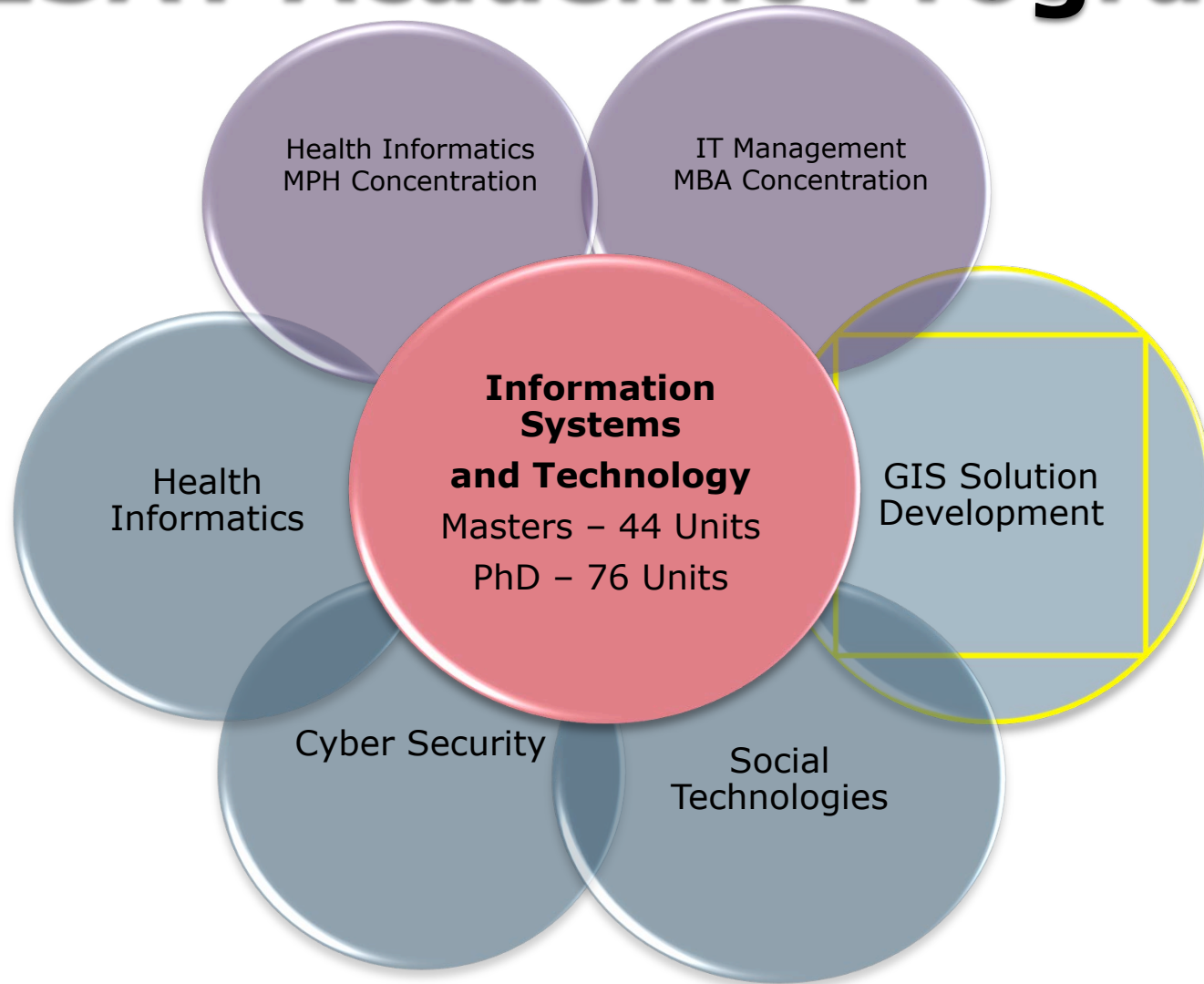
Geographic (Spatial) information science: is the science behind the technology and considers fundamental questions raised by the use of systems and technologies; GIS-T is the science needed to keep technology at the cutting edge; it is a *transdisciplinary* field with many disciplines contributing to these issues.

Wave 1: geography, cartography, urban planning,

Wave 2: information/computer science, social sciences, economic and business applications.

(Adapted from Goodchild, 2002)

CISAT Academic Programs



A Brief History of GIS and Esri at CGU

- 1998 - Spatial Cognition of Neighborhoods Study
- 2000 - GIS Course Offered at CISAT
- 2003 - First Spatial Information System Development Dissertation
- 2006 - GIS Concentration Started
- 2008 - Named one of three inaugural EDC's by Esri
- 2009 - SafeRoadMaps achieves 10 million visits
- 2010 - First Esri MOU signed
- 2011 - CISAT Refines Concentrations, Accelerates GIS
- 2012 - Advanced GIS Lab formed to Enhanced CGU Research
- 2013 - Renewed Esri MOU and Scholarship Program
- 2013 - Horan and Hilton Invited to White House
- 2013 - GIS Becomes Top CISAT Concentration
- 2013 - CISAT Expands GIS Efforts Across CGU
- 2013 - Esri Visit for GIS Day and Student GIS Club Formed
- 2014 - Board of Trustees Approve Additional GIS Positions

The CGU Focus: Video



Location Analytics at CGU

<http://is.cgu.edu>

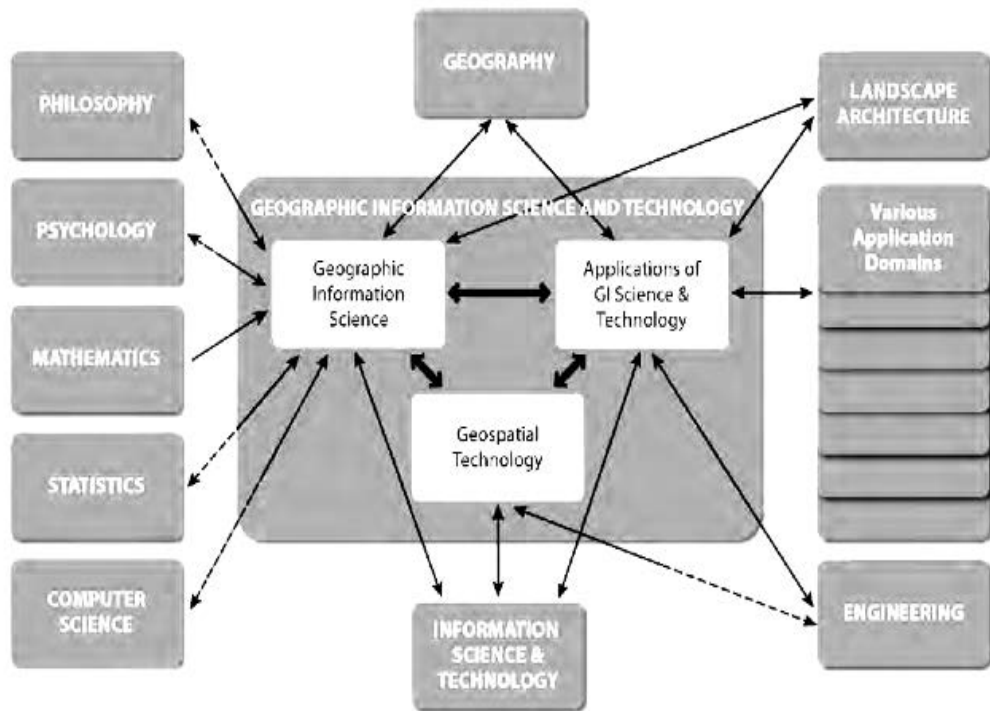
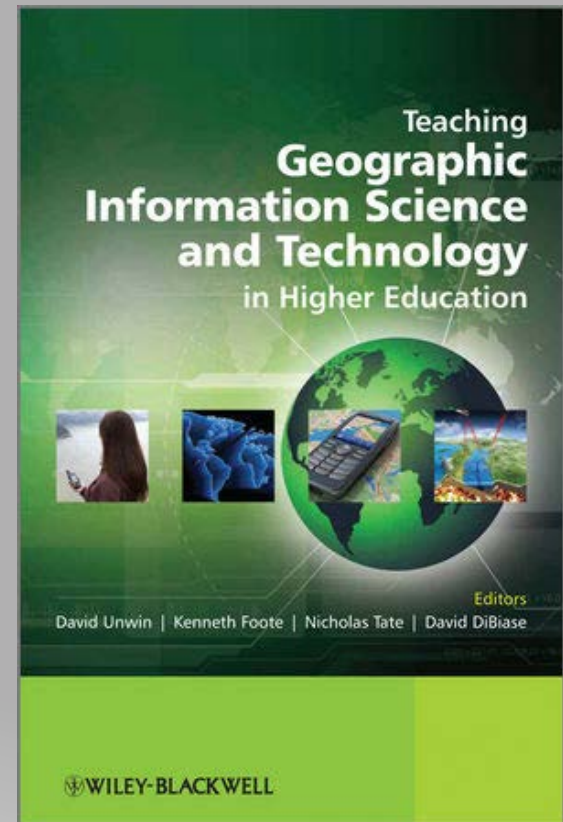


Figure 1: The three sub-domains comprising the GIS&T domain, in relation to allied fields. Two-way relations that are half-dashed represent asymmetrical contributions between allied fields.



Practicum: Practice-Based Learning



The Advanced GIS Lab in CISAT focuses on advanced location analysis and the research and development of advanced GIS solutions. The Lab stresses an integrated approach to knowledge, research, and problem-solving that takes the core ideas, methods, and concepts from several disciplines and uses them to study a broad range of problems.

Esri Partnership and Development Center (EDC)

One of the founding Esri Development Centers, (EDC).

New MOU supports education, research and training collaboration.

Applied focus unites research on pressing issues with new platforms, tools, and techniques



NEWS & EVENTS

OFFICE OF COMMUNICATIONS

[PROSPECTIVE STUDENTS](#) | [CURRENT STUDENTS](#) | [MYCAMPUS](#) | [ALUMNI & FRIENDS](#) | [CONTACT US](#)


[ABOUT CGU](#)

[NEWSROOM](#)

[FLAME ONLINE](#)

[EVENTS CALENDAR](#)

[FACULTY EXPERTS](#)

 [Subscribe to News RSS](#)

Claremont Graduate University announces new partnership with Esri

Tuesday, June 04, 2013

Claremont Graduate University (CGU) has entered into a new comprehensive partnership agreement with Esri, the world's leading supplier of geographic information system (GIS) software.

Under the three-year agreement, Esri and CGU will provide scholarships for at least three Esri employees to begin a CGU master's or doctoral degree program per year. Additionally, Esri and CGU will engage in collaborative research and initiate a clinic program utilizing resources from several CGU programs, including the Center for Information Systems and Technology (CISAT), the Peter F. Drucker and Masatoshi Ito Graduate School of Management, and the Institute of Mathematical Sciences.

GIS Concentration Highlights

Foundation Courses (Required)

IST 370: Geographic Information Systems: Essential Concepts

IST 371: GIS Solution Development

Advanced Courses (Choose 1)

IST 372: Advanced GIS Analytics and Solution Development

IST 380: Geospatial Business Intelligence

IST 380: Health Geoinformatics

IST 380: Spatial Thinking and Statistics (planned)

Specialized Topics (Choose 1)

IST 380: Cloud Computing

IST 380: Mobile Applications

Practicum (Required)

IST 373: GIS Practicum

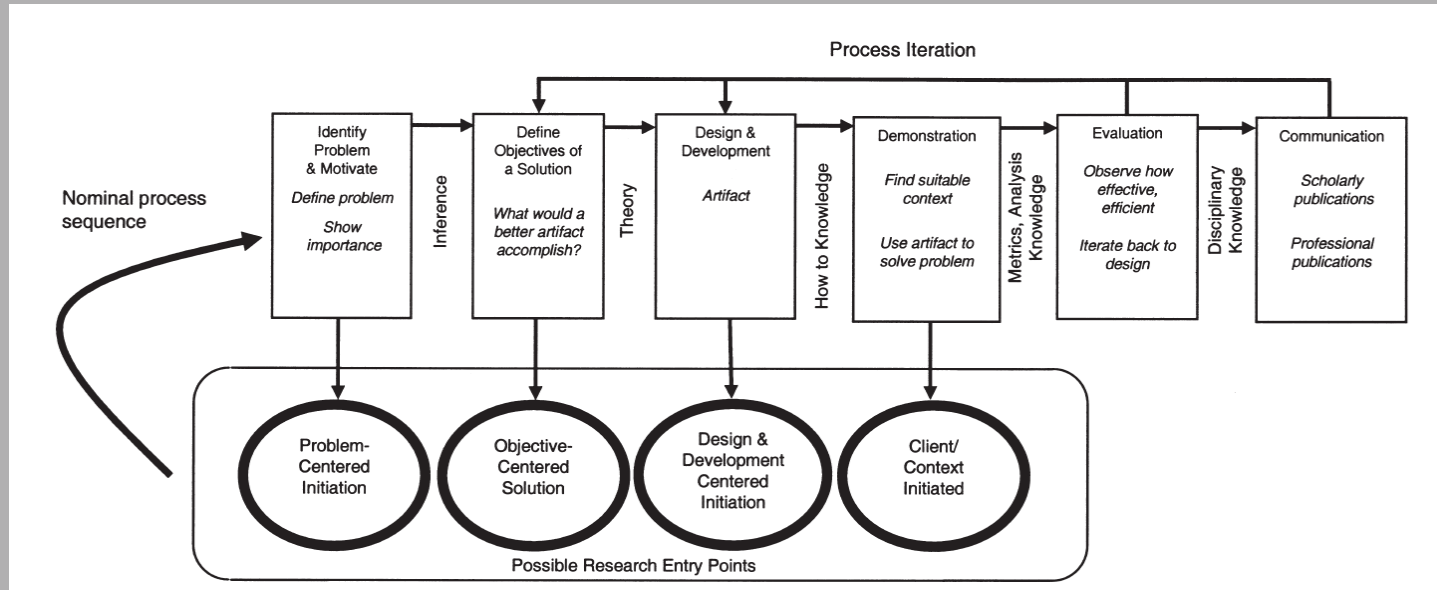
360 student enrollees in GIS
Concentration Courses, 204 since
2011



Current Lab Projects Underway

Topic and/or Organization	PI	GIS Lab Role	Students	Domain
Freight Rail BNSF / Esri	Horan / Hilton	Lead	Yaser Khouja, Ziyun Xu (Riki), Watanyoo Suksa-Ngiam	Transportation
Tribal Transportation Safety / Esri	Horan	Lead	Ziyun Xu (Riki), Rachel Camacho	Transportation
Health Spatial Decisions Inland Empire / Esri	Horan / Hilton	Lead	Michael McElroy	Health
Spatial De-identification / Esri	Horan	Lead/Support	Abdullah Murad	Healthcare/Statistics
Community Vital Signs Data	Hilton	Lead	April Moreno	Health
Tobacco Use Among Youth	Hilton	Lead/Support	Hilary Tannenbaum, Rachel Camacho	Health
Personal Safety Spatial Decision Making	Horan / Hilton	Lead	Juliette Gutierrez	Health
Retail Food Environment Index	Hilton	Lead	Hilary Tannenbaum, Rachel Camacho	Health
Montana Native American Voting	Jean Schroedel	Support	Ziyun Xu (Riki)	Political Science
Energy Infrastructure	Hal Nelson	Support	John Urata	Political Science
Community College Engagement	Cecilia Rios-Aguilar	Support	Maria Reyes	Education
GIS in High School Teaching - CHS	Hilton	Lead	Sarah Osailan	Education
Upland USD Spatial Data Analysis	Hilton	Lead/Support	Rachel Camacho, Marco Antonio Cruz	Education
Las Vegas Real-time Spatial Social Analysis	Horan / Hilton / Zak	Lead/Support	Sarah Osailan, Yaser Khouja, Omar Aboulola	Economics

Design Science Research Methodology



Design Science is an outcome based information technology research methodology, which offers specific guidelines for evaluation and iteration within research projects.

Design science research requires the creation of an innovative, purposeful artifact for a special problem domain.

Design Science Research Methodology

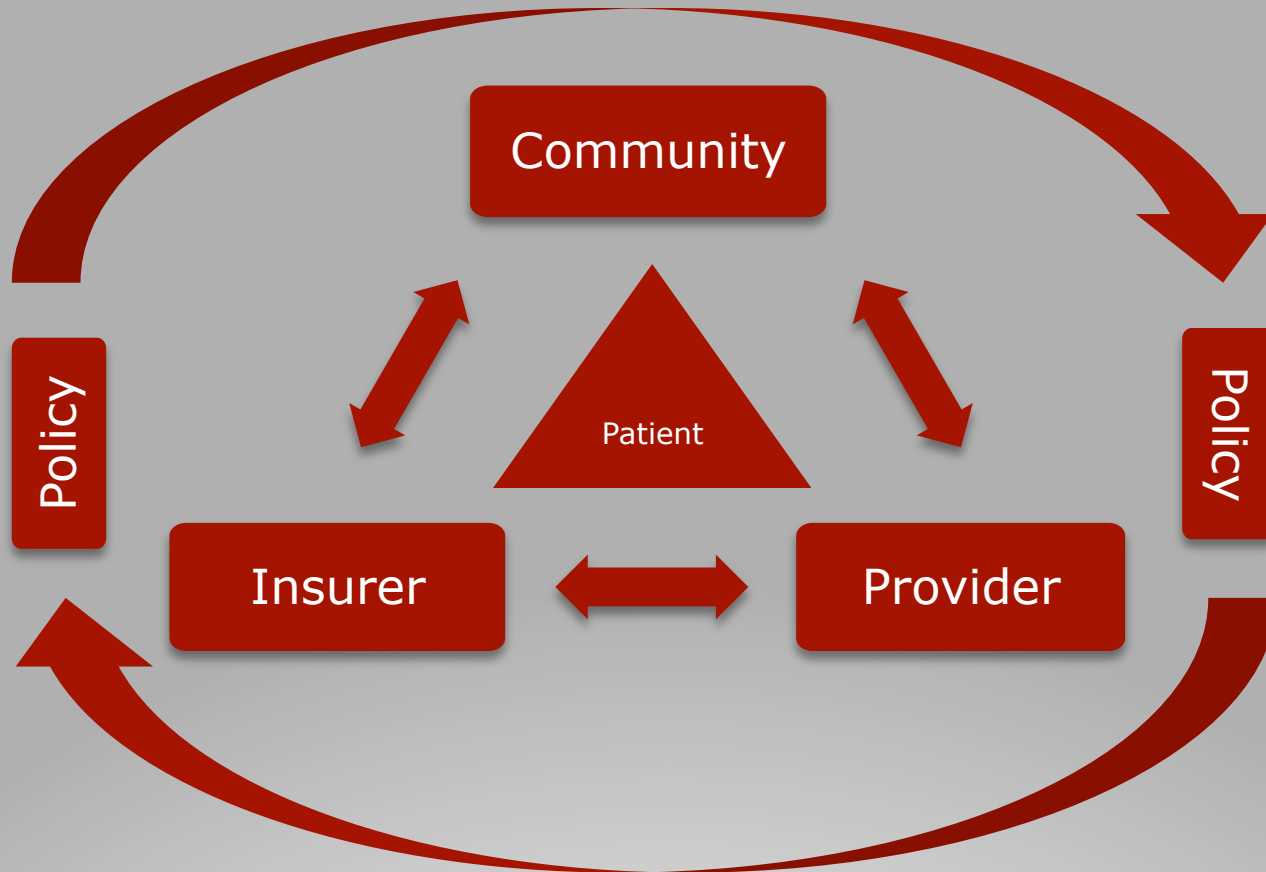
Lab Research and Development Methodology

The artifact must be evaluated in order to ensure its utility for the specified problem. In order to form a novel research contribution, the artifact must either solve a problem that has not yet been solved, or provide a more effective solution.

Both the construction and evaluation of the artifact must be done rigorously, and the results of the research must be accepted from both presented effectively both to technology-oriented and management-oriented audiences.



GeoHealth Informatics Ecosystem



GIS as a Business Intelligence Tool

Michael McElroy

Gartner: 8 essential components of BI

- Reporting
- Dashboards
- Ad hoc query
- Search-based
 - OLAP
- Interactive visualization
 - Scorecards
- Predictive modeling
 - Data mining

GIS is similar in that it:

- Provides highly robust capabilities in most of these areas, and extends visualization and predictive modeling beyond traditional BI systems
- Allows for the leveraging of the spatial component of data, which is not a native component of traditional BI systems
- Maps and other visualization can help tell a story that is otherwise difficult to realize in traditional scorecards, charts, tables, etc.

Research Questions

How can spatially enabled data inform healthcare decision making at the community and clinical service delivery level?

What current healthcare business imperatives could benefit from a spatial perspective?

What are the range of organizational decisions that might be affected by introducing GIS-based IT artifacts?



Methodology

Action Design Research

Model specific business processes that have an impact on decision making

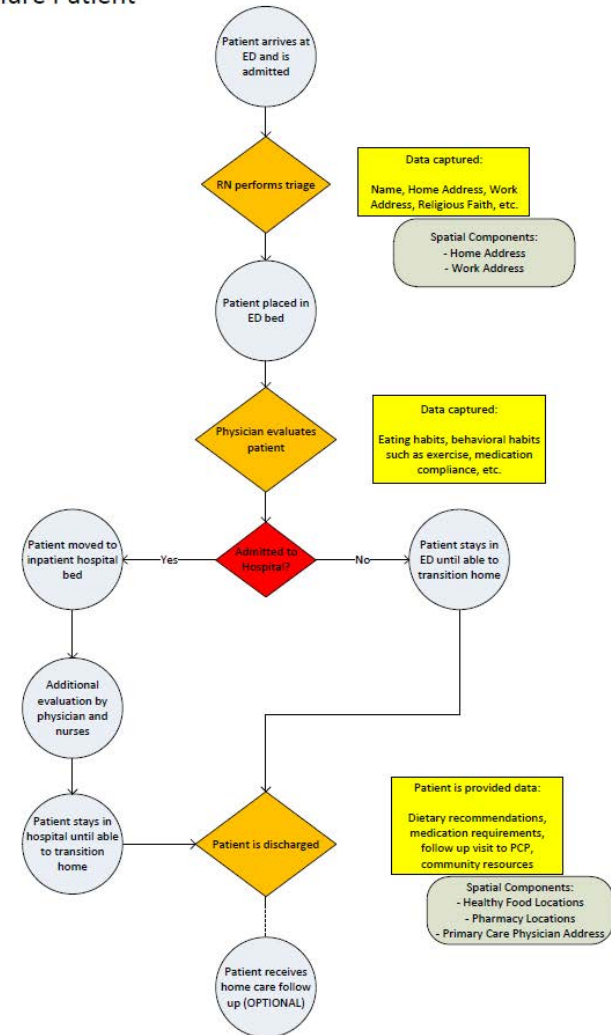
Use the identified processes and spatial data to build IT artifacts which showcase how geospatial technology can improve decision making

Extract, from those processes, datasets that have a spatial component

Process Modeling

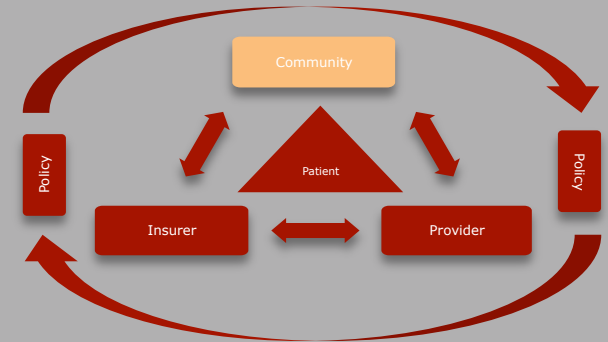
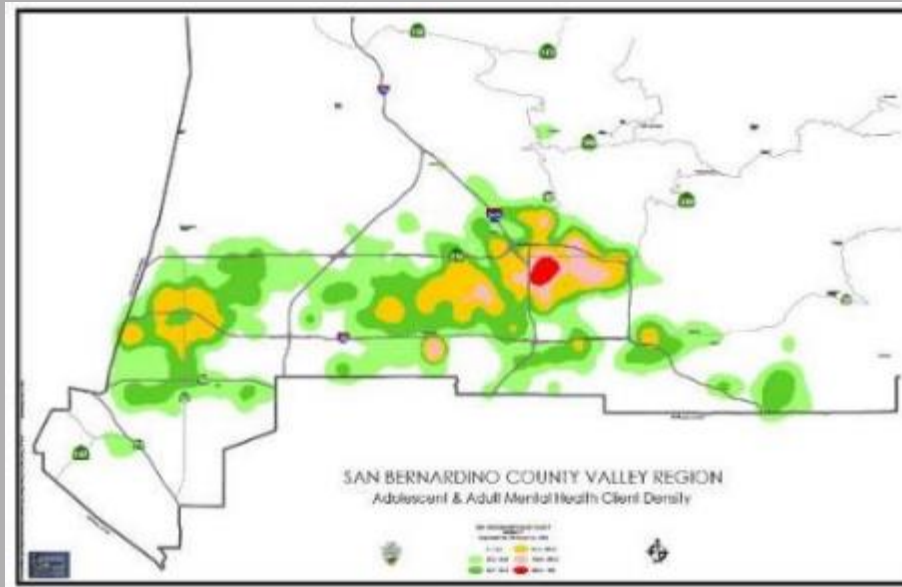
- Identify business imperatives
- Model current process
- Identify spatial components
- Define opportunities for use of GIS
- Prototype application

Basic Admission Workflow:
Heart Failure Patient



Behavioral Health Assessment

San Bernardino County Behavioral Health

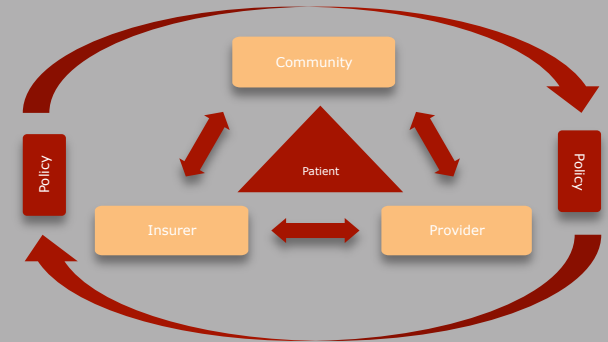


- Visualization of mental health client density
- Overlay existing resources
- Useful to plan where to place new resources

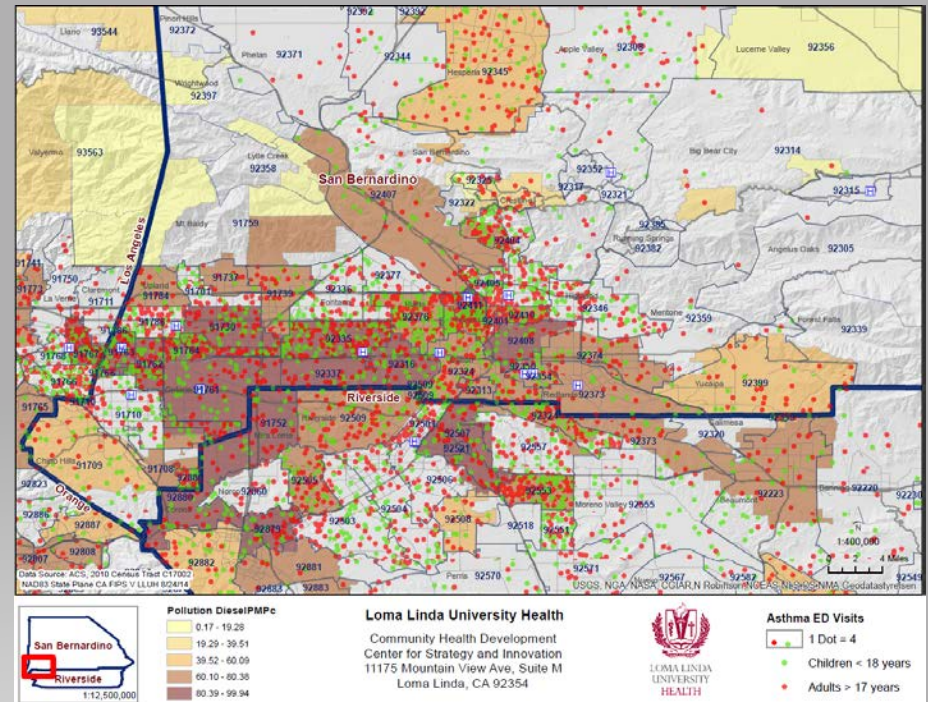


Childhood Asthma Risk

Loma Linda University Health (LLUH)



- Map environmental factors in real time
 - Wind
 - Air Quality
- Overlay location of children with history of asthma
- Develop a risk score for potential attack
- Send targeted notifications
 - Parents
 - School Nurses



Spatial Analysis of Freight Economy and Supply Chain



GeoSpatial Supply Chain: Freight Rail

- The purpose of this study was to focus the nation's attention on the economic role of freight system in the US.
- The focus of the study is on industry clusters in Minnesota and the relationship of these geographic economic clusters on the freight rail system.
- The study has interwoven research and outreach elements.



Study Methods

- **Economic and Related GIS Data Analysis**
 - Industry cluster, GSP shift-share, and GIS analysis
- **National and Regional Expert Interviews**
 - Class 1 and shortline railroads; federal, state and local policymakers; system users; economic development officials.
- **Report Reviews and Synthesis**
 - Various sources: NCFRP, SHRP, economic research, Minnesota state freight and rail plans.
- **Outreach and Feedback**
 - Interim/ Final Presentations, TPEC Advisory Committee.



Minnesota Industry Cluster Story Maps



Metal Mining Industry Cluster

Iron ore has been one of Minnesota's most abundant natural resources for many years...

[Click to open story map](#)



Agriculture Industry Cluster

Agricultural products have been a staple of the Minnesota economy...

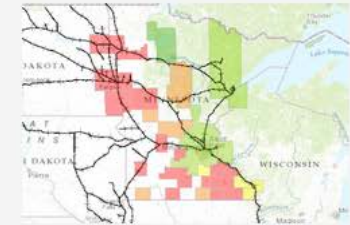
[Click to open story map](#)



Food Processing Industry Cluster

The food processing industry of Minnesota is one of the most competitive industries...

[Click to open story map](#)



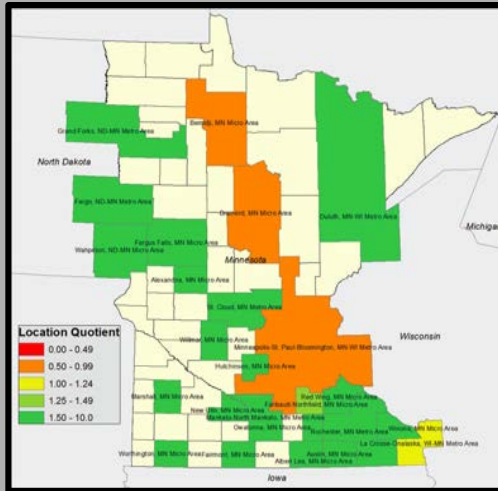
Interactive Atlas

Visit our interactive map, loaded with data about freight rail's impact... (high bandwidth required)

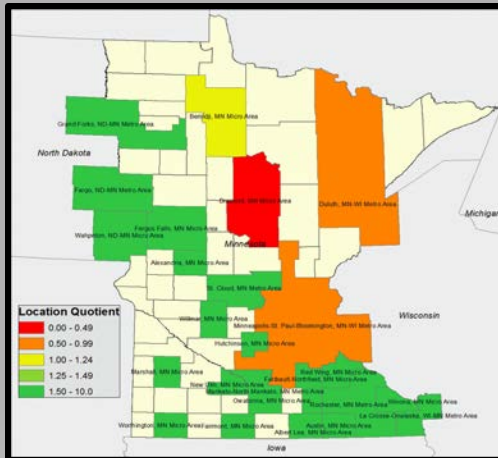
[Click to open interactive atlas](#)

<http://freighteconomy.org>

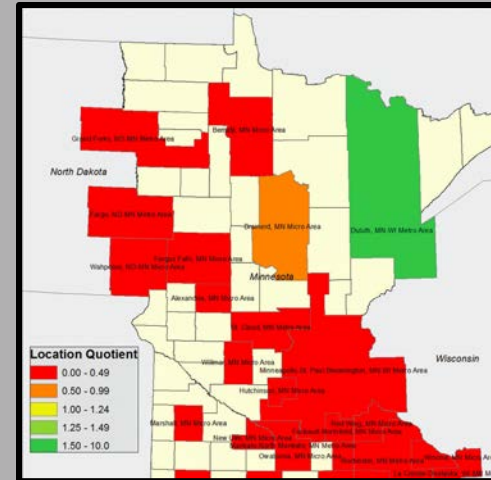
Agricultural Products



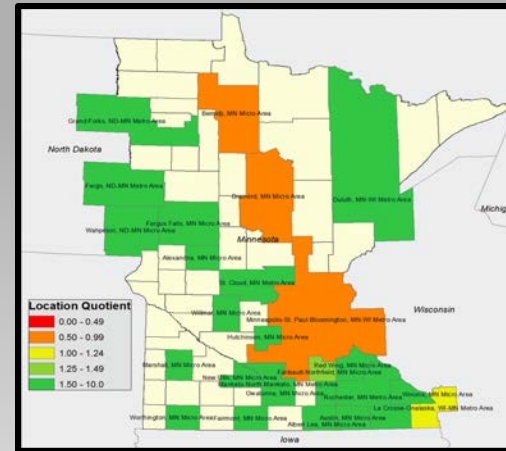
Processed Food



Metal Mining



Heavy Machinery

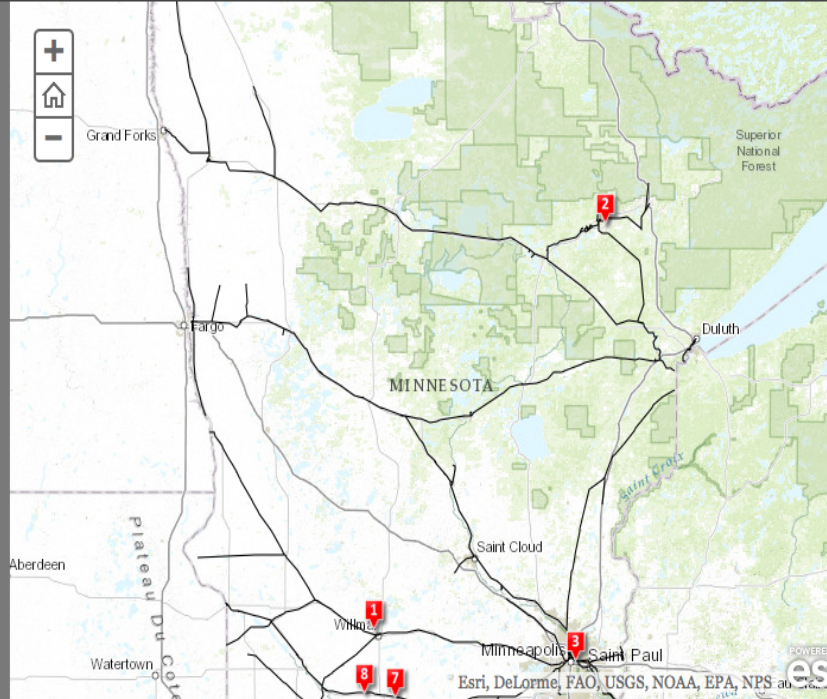


"The Best Single Map.."



Understanding and Enhancing the Value of Freight Economy in Minnesota

Freight rail is important to Minnesota's economic competitiveness and plays a vital role in key Minnesotan industries. State GDP has been higher than the national average in a few key areas - some of which are directly linked to freight.



1 Willmar Poultry Company (WPC)



2 Essar Steel Minnesota LLC



3 Murphy Warehouse



4 Minnesota Prairie Line: An example of the importance



5 Seneca Foods Plant - Arlington MN (Population



6 Heartland Corn Products

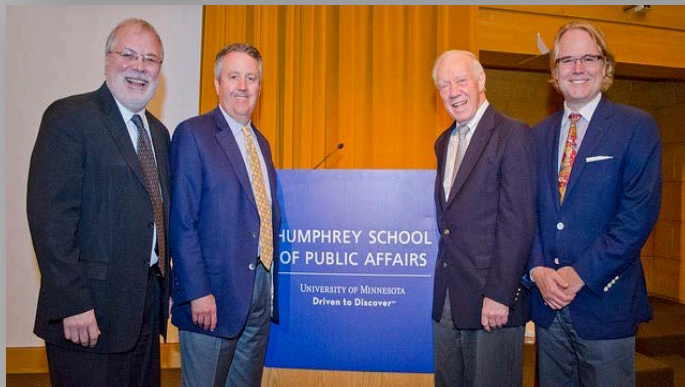
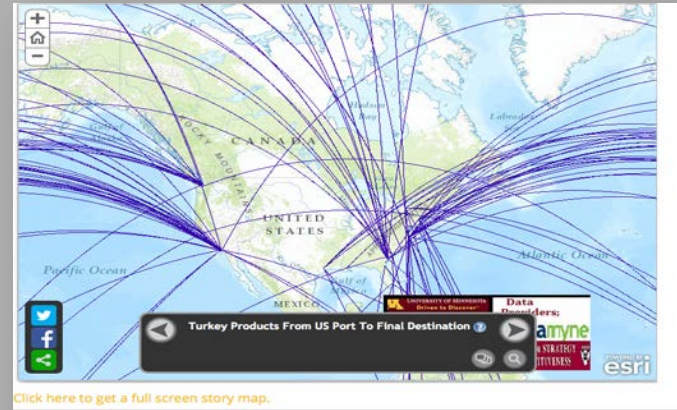


7 Bird Island Bean Company



8 Southern Minnesota Beet Sugar Cooperative

Impact



A screenshot of a news article from KSTP.com. The article is titled "New Study: Freight Rail Linked to Healthier Economy" and is dated "Sep. 19, 2013, 10:17 PM CDT [2:38]". The article is updated on "09/19/2013 11:20 PM" and created on "09/19/2013 10:05 PM". The author is "Naomi Pescovitz". The article includes a video player showing a freight train and a play button. The article is from "5 EYEWITNESS abc NEWS".

New Directions



Big Data Offerings @ CISAT/CGU



Sentiment Analysis for Business: A Student Case Study

Sean Bjurstrom

Claremont Graduate University, Center for Information Systems and Technology

Introduction

What is it?
Sentiment Analysis is the use of natural language processing and machine learning tools to analyze unstructured data such as text to determine the key concepts and sentiment of that text.

Why use it?
Sentiment analysis can help an organization gain better insights as to how the public views its organization. It can be used to determine if an organizational campaign is having an impact and in what way. It can help answer such questions as product and how do they view it? Sentiment analysis has been used in industries such as healthcare, politics, marketing, and public relations.

Objectives

- Understand the customer's image of the organization and products being reviewed.
- Collect "real" reviews of our marketplace and determine how the public views these organizations. Could potentially be a better source of information than those collected by a dealership after a purchase has been made.
- Create a dashboard that could be used by an organization to support decision making.
- Develop a process for performing sentiment analysis using open source, big data tools.
- Process models to be easily available to handle massive datasets.

Methods

Methodology

Collect Social Media

- Extract reviews and data about reviews from sources such as websites and Twitter.
- Store all data in an AWS S3.

Conduct Sentiment Analysis

- Perform sentiment analysis of text. Extract keywords, entities, and review sentiment.
- Class data.
- Normalize scores.

Visualize

- Import into Tableau for visualization and knowledge discovery.

1. Collect Social Media 2. Conduct Sentiment Analysis 3. Visualize

Open Source, Big Data Methodology

The methodology for performing sentiment analysis using open source, big data tools is very similar to that of performing the sentiment analysis on a desktop. The only difference is in Step 2. In Step 2, instead of using Python with Akshara API to perform the analysis, Amazon's Web Services' big data platform is used. The analysis is performed in three stages:

- Create and run job using Amazon ElastiMapReduce.
- Split job among Amazon EC2 servers to perform analysis in parallel.
- Combine results from each server to receive overall results.

Why use big data tools?

The first method works well for several thousand reviews. If 10,000 or 100,000 reviews need to be analyzed, it will take far too long to perform. Method 2, is infinitely scalable by simply adding more servers and the processing time to perform 10,000 analyses on 100,000 will be similar.

Definitions

- EC2: Amazon's Elastic Cloud Compute, a cloud server hosting platform.
- S3: Amazon's massive file storage service.
- Elastic MapReduce: A service that is part of Amazon's Hadoop platform.

Open Source, Big Data Methodology

1. Collect Social Media 2. Conduct Sentiment Analysis 3. Visualize

Results

Average Sentiment Score of Social Media Sites

Average Sentiment Score by Social Media Site

1. Collect Social Media 2. Conduct Sentiment Analysis 3. Visualize

Conclusions

Big data sentiment analysis can be done without a semester using commercial (e.g. Amazon) and related open source tools. This provides revealing insights for business, policy, or research. Commercial providers such as SPIN are very useful in quickly identifying trends, making it roughly equivalent to manual analysis in complex situations.

In short it's a new tool that will be common place in the analytical arsenal.

CGUAT is dedicating resources to develop this student research capacity at CGU!

References

- Akshara API: <http://www.aksharaapi.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>
- Amazon.com: <http://www.amazon.com/>

Transdisciplinary Big Data Fundamentals

BIG DATA, NEW WORLDS: DATA ANALYTICAL TOOLS ACROSS THE DISCIPLINES

TNDY 4051 (SPONSORED BY T-STUDIES PROGRAM, IN COLLABORATION WITH H. TCHALIAN, DRUCKER SCHOOL)
This course focuses on using Big Data tools and technologies effectively across various disciplines and settings – social sciences, humanities, information systems, business, policy, and healthcare. The enormous volume, velocity, and variety of data created every day from social media in the form of Twitter feeds, Facebook posts, and countless other platforms and sensors shows no signs of abating. The class takes a wide-ranging, transdisciplinary, hands-on approach to introducing, learning, using, and managing such Big Data opportunities. In the course, you not only explore the what, but the how and the why, of big data analysis in different disciplines.

Data Science Analytics

DATA SCIENCE PROGRAMMING –

IST 380 (SPONSORED BY BLAIS GRANT, IN COLLABORATION WITH Z. DODDS, HARVEY MUDD COLLEGE)

The term "Data Science" is a recent one that captures gamut of analytic approaches to data gathering, analysis, visualization and utilization. This course provides you with an introduction into computational thinking, and how this thinking guides software programming for Data Science applications. You will be taken through a hands-on and engaging process for learning and using Python and related Data Science programming languages.

KNOWLEDGE DISCOVERY AND DATA MINING – IST 333

Do you want to help predict the next popular product or customer-product matching? Find what will trend next on social media? Help prevent crimes? Then the Knowledge Discovery and Data Mining course can teach you the data mining techniques to do that. In this course you will look at the entire knowledge discovery process, including interesting data mining techniques for Big Data that can be used in a wide variety of settings (business, science, web).

GEOSPATIAL SOLUTIONS – IST 371

Does spatial location matter in your research or practice? This course introduces you to the design and development of geospatial (GIS) applications and systems. You will develop the fundamental components of a GIS solution and demonstrate your mastery of basic GIS solution development practices with the delivery of a prototype GIS. The course ensures that you are exposed to the most current locational analytic tools for use in the emerging geospatial Big Data world.

Sample Domain Applications

BUSINESS: TECHNOLOGICAL INNOVATION AND ENTREPRENEURSHIP – IST 316

This class is designed to expose you to the transformative and wealth generating effect innovation has on enterprise in our increasing globalized and data rich society. There are several key themes you will discuss during the course, such as: innovation is the heart of a business and the foundation for entrepreneurship, current trends in technology (including Big Data) provide a new set of challenges and opportunities for innovation, and successful enterprises organize themselves to continuously innovate throughout these technological changes.

HEALTH: GEOINFORMATICS - IST 380C (Summer 2015)

Did you know that where you live can have a strong association with your health? The use of geospatial technologies and information can improve our understanding of the relationships between, people, location, time, and health and healthcare issues. In this course you will examine how these technologies and Big Data analyses can assist in: discovering and eliminating disease, disease prevention and health promotion for community health, and healthcare service planning and delivery.



Toward Predictive Analysis via Social Media, Big Data, and GIS

Anthony J. Corso

Claremont Graduate University

Introduction

To support dissertation research I had become crime, the linguistic analysis of social media, and data of public domain needed to be observed. Among "Naïve Bayes", "Support Vector Machines", and "The Spa" Analytic tools covered the traditional black-box solution for this type of research and dimensional integration systems of data integration. Outcomes, unexpectedly, appeared linear research outcomes and will augment GIS, wishing to integrate social media corpora in predictive ways via GPUs.

Objectives

Design the small amount of data given in a social media microtext, i.e. a tweet, if its content is properly analyzed it can be used as input for a GIS model where predictive capabilities are achieved. For example:

- Recognition of latent events in order to record its observation by an occurrence, or
- Prediction of real-time trends via social media and behavior of historical record.

Natural Language Processing of Social Media

Social Media Vocabulary Decay

Social Programs vs. Income

Social Media vs. Income

Crime vs. Income

Collective Data

Discussion

In order to support a social media hot spot analysis GIS artifact, ostensibly allowing predictive crime-based social media research, a number of issues must be considered and resolved. Although ad hoc Big Data analysis has been successful it is largely represented via single domain inquiry. The observation of multi-domain dataset processing needs significant consideration in conclusion can be drawn via integration of disparate data. In addition, an artifact's framework needs to support data analysis including event statistical techniques, e.g. Laplace smoothing and logistic regression of constituent data streams. This will enable large-scale dataset evaluation to be conducted in a meaningful but iterative way. When fully implemented the former will create a robust artifact that enables GIS geospatial and spatial correlation processes, so further consider, extend, and uncover evident data relationships previous research omits because of shortcomings with respect to single-domain datasets. The figure below illustrates initial outcomes for such an artifact.

Conclusion

Although integrating public domain and crime data are not novel, combining them with ad hoc social media data streams in predictive ways via GIS is. In this brief pilot study on incident-based crime data and data of public domain a positive correlation among the data was established. The work specifies the framework needed for future social media GIS projects.

References

- Agarwal, A., Xia, B., Vovsha, I., Barnrow, O., & Pramanick, R. (2011). Sentiment analysis of Twitter data. Paper presented at the Proceedings of the Workshop on Languages in Social Media, Portland, Oregon.
- Barlow, L., & Fong, J. (2010). Robust conditional distributions on Twitter from biased and noisy data. Paper presented at the Proceedings of the 23rd International Conference on Computational Linguistics, Poster, Beijing, China.
- Boatman, P. J., & Huntington, P. L. (1981). Environmental cartography. Sage Publications: Beverly Hills, CA.
- Caplan, J. M., & Kennedy, L. W. (2010). Risk Terrain Modeling Manual. Newark, NJ: Rutgers Center on Public Security.
- Caplan, J. M., & Kennedy, L. W., & Miller, J. (2010). Risk Terrain Modeling: Understanding "Criminality" Through GIS Methods. *Crime Prevention and Justice Quarterly*, 28(2), 369-381. doi: 10.1080/07418825.2010.486037
- Ko, A., Hwang, R., & Hwang, K. (2009). Twitter sentiment classification using latent Dirichlet allocation. *CS234N Project Report*, Stanford, 1-12.
- Kennedy, L. W., Caplan, J. M., & Pina, E. (2011). Risk clusters, hotspots, and spatial neighborhood risk terrain modeling as an algorithm for police tactical allocation strategies. *Journal of Quantitative Criminology*, 27(3), 139-162.
- Mancoff, M. (2013). Data Just Right: Introduction to Large-Scale Data and Analytics. Pearson Education.
- Pina, E. L. (2012). Using Positive and Negative Binomial Regression Models to Measure the Influence of Risk on Crime Incident Counts. *Hawthill, J. H. (2004). The hotspot matrix: A framework for the operational management of crime in public places. Practice and Research*, 5(1), 5-23.
- Saif, H., Ha, Y., & Akmal, H. (2012). Semantics-oriented analysis of twitter. *The Semantic Web: ISWC 2012* (pp. 698-724). Springer.
- Sanz, P., Mito, M. W., Richards, S. C., Bhattacharjee, T., & Malambo, L. (2012). Social Media Analysis and Geospatial Crime Report Charting for Crime Prediction & Prevention.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques*. Elsevier.

Contact

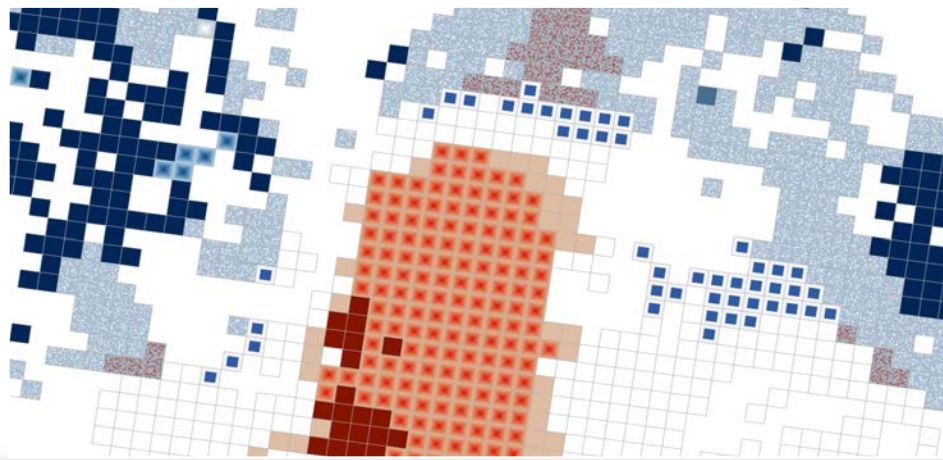
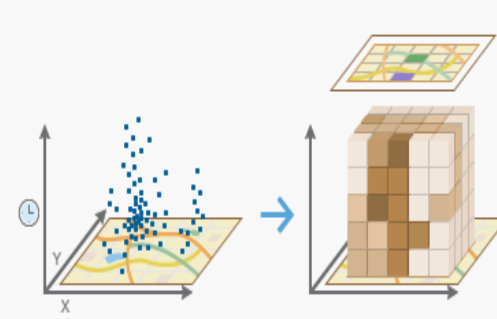
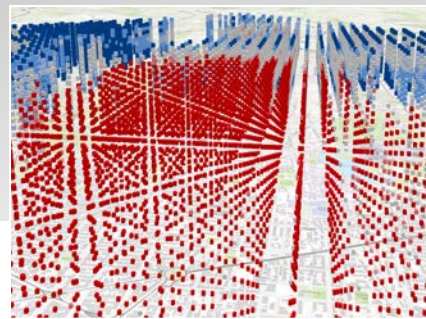
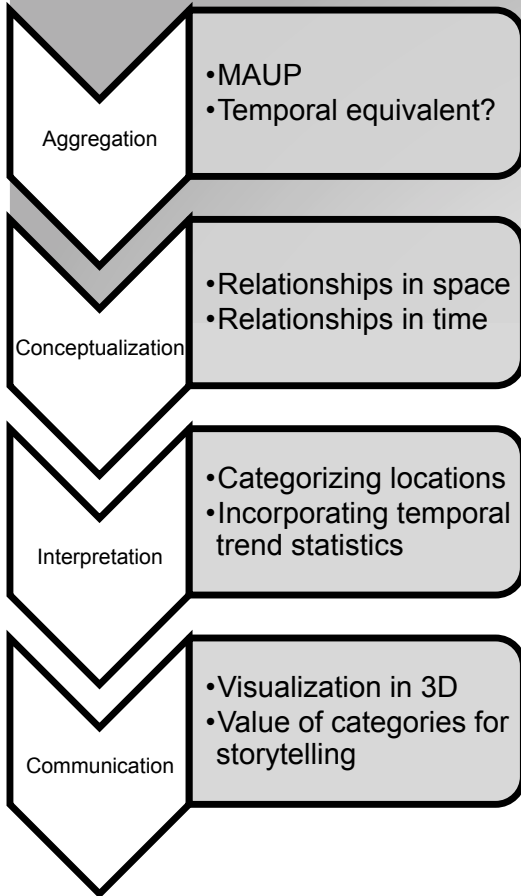
Mr. Corso is a Ph.D. candidate pursuing a degree in Information Systems from Claremont Graduate University. He joined California Baptist University in 2007 and is an Assistant Professor in the College of Engineering.

ajcorso@calbaptist.edu



Emerging Hot Spot Analysis

A Space-Time Approach to Pattern Mining
 Lauren Bennett



- New Hot Spot
- Consecutive Hot Spot
- Intensifying Hot Spot
- Persistent Hot Spot
- Diminishing Hot Spot
- Sporadic Hot Spot
- Oscillating Hot Spot
- Historical Hot Spot
- New Cold Spot
- Consecutive Cold Spot
- Intensifying Cold Spot
- Persistent Cold Spot
- Diminishing Cold Spot
- Sporadic Cold Spot
- Oscillating Cold Spot
- Historical Cold Spot
- No trend detected
- < all other values >

For More Information

Acknowledgements



Brian Hilton
Mike McElroy
Lee Munnich
Yaser Khouja
Lauren Bennett
Anthony Corso
Sean Bjurstorm

Tom.Horan@cgu.edu
<http://is.cgu.edu>