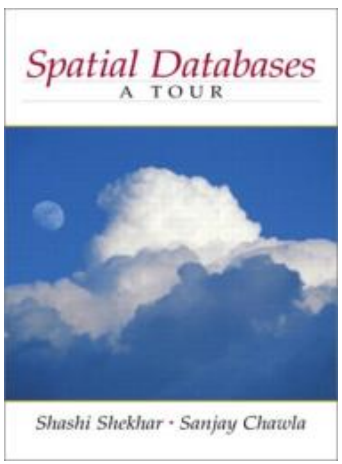
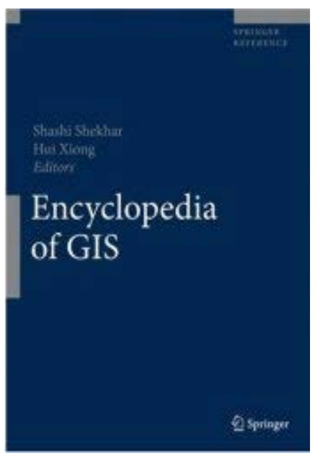
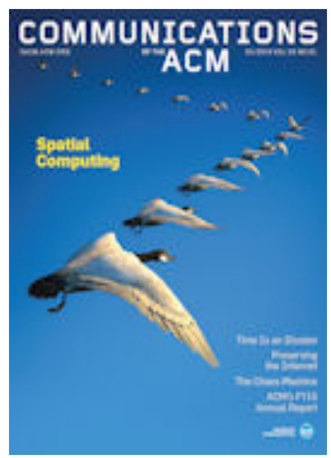
From GPS, Google Maps and Uber to Spatial Computing

December 2019

Shashi Shekhar
McKnight Distinguished University Professor
Computer Sc. and Eng., University of Minnesota
www.cs.umn.edu/~shekhar



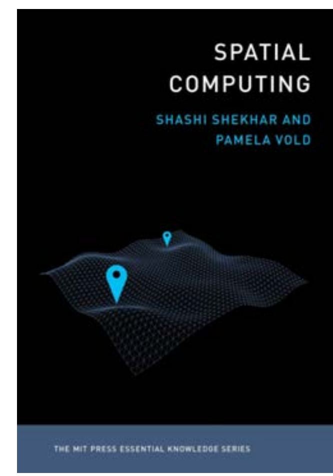





University Consortium for Geographic Information Science

A UCGIS Call to Action:
Bringing the Geospatial Perspective to Data Science Degrees and Curricula

Summer 2018



Thanks: Sponsors: NSF, USDOD, NASA, USDOT, USDOE, Ford URP, IBM, Microsoft, …
UMN: CTS, IonE, DTC, Spatial Computing Group students

Spatial Computing: Recent Examples

ESRI ArcGIS X

SaTScan™
Software for the spatial, temporal, and space-time scan statistics

OGC™
Open Geospatial Consortium, Inc.

ORACLE SPATIAL

IBM Smarter Planet

vs

Google Earth

waze
Outsmarting Traffic, Together.

OpenStreetMap

Ushahidi

Leading Market Players

LimeBike SPIN

Lime Spin Bird

Google Earth Engine

MONSANTO

THE CLIMATE CORPORATION

The Changing World of Spatial Computing

	Last Century	Last Decade
Map User	Well-trained few	Billions
Mappers	Well-trained few	Billions
Software, Hardware	Few layers, e.g., Applications: Arc/GIS, Databases: SQL3/OGIS	Almost all layers
User Expectations & Risks	Modest	Many use-case & Geo-privacy concerns

It is widely used by Government!

Q? Which agencies sowed seeds for Google Maps?

May 18th, 2011

Congressional Research Service

Table 1. Members of the Federal Geographic Data Committee (FGDC)

Dept. of Agriculture	Environmental Protection Agency
Dept. of Commerce	Federal Emergency Management Agency
Dept. of Defense	General Services Administration
Dept. of Energy	Library of Congress
Dept. of Health and Human Services	National Aeronautics and Space Administration
Dept. of Housing and Urban Development	National Archives and Records Administration
Dept. of the Interior (Chair)	National Science Foundation
Dept. of Justice	Tennessee Valley Authority
Dept. of State	
Dept. of Transportation	Office of Management and Budget (Co-Chair)

Source: Peter Folger, Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges. Congressional Research Service. June 8th, 2009.

Deconstructing Precision Agriculture

#AgInnovates2015

Wednesday, March 4, 2015
Reception | 5:00 to 7:00 pm
House Agriculture Committee Room,
1300 Longworth House Office Building,
Washington, DC

Think Moon landing.

Think Internet.

Think iPhone and Google.

Think bigger.

Come hear U.S. farmers, leading agriculture technology companies, and scientists tell how hey work together to fuel U.S. innovation and the economy to solve this global challenge.

The event will exhibit three essential technologies of precision agriculture that originated from a broad spectrum of federally funded science: Guidance Systems and GPS, Data & Mapping with GIS, and Sensors & Robotics.

Moderator

Raj Khosla, Professor of Precision Agriculture at Colorado State Univ.

Farmers

David Hula, of Renwood Farms in Jamestown, Virginia

Rod Weimer, of Fagerberg Produce in Eaton, Colorado

Del Unger, of Del Unger Farms near Carlisle, Indiana

Speakers

Mark Harrington, Vice President of Trimble

Carl J. Williams, Chief of the Quantum Measurement Division at NIST

Bill Raun, Professor at Oklahoma State Univ.

Marvin Stone, Emeritus Professor at Oklahoma State Univ.

J. Alex Thomasson, Professor at Texas A&M Univ.

Dave Gebhardt, Director of Data and Technology at Land O'Lakes/WinField

Shashi Shekhar, Professor at the Univ. of Minnesota

RSVP

http://bit.ly/1CoOYoa

Hosted by the Congressional Soils Caucus

In partnership with

Agricultural Retailers Association
American Society of Plant Biologists
American Physical Society
American Society of Agronomy
Association of Equipment Manufacturers
Coalition for the Advancement of Precision Agriculture
Computing Research Association
CropLife America
Crop Science Society of America
PrecisionAg Institute
Soil Science Society of America
Task Force on American Innovation
Texas A&M AgriLife
Trimble
WinField

This is about feeding the world.

Economy & Spatial Computing

McKinsey Global Institute

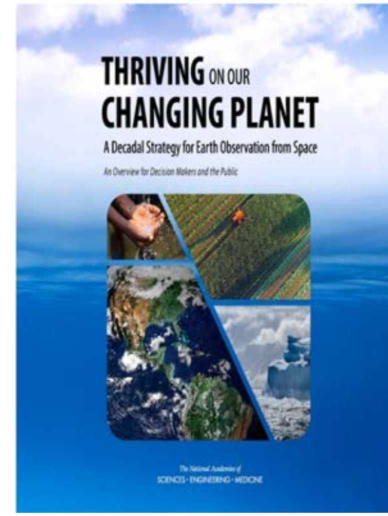
The study estimates that the use of personal location data could save consumers worldwide more than $600 billion annually by 2020. Computers determine users' whereabouts by tracking their mobile devices, like cellphones.

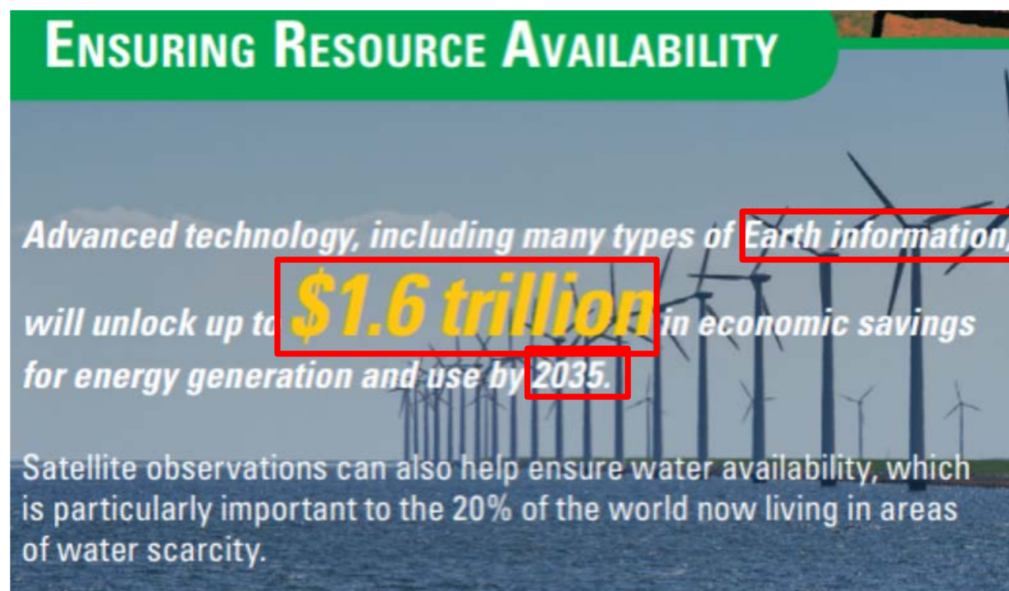
The biggest single consumer benefit, the study says, is going to come from time and fuel savings from location-based services tapping into real-time traffic and weather data that help drivers avoid congestion and suggest alternative routes. The location tracking, McKinsey says, will work either from drivers' mobile phones or GPS systems in cars.

The New York Times

Published: May 13, 2011

Big data: The next frontier for innovation, competition, and productivity





CCC Visioning Workshop: Making a Case for Spatial Computing 2020

http://cra.org/ccc/spatial_computing.php

Computing Community Consortium

We support the computing research community in creating compelling research visions and the mechanisms to realize these visions.

HOME | ABOUT | YOUR VISION | ACTIVITIES | RESOURCES | CONTACT | GO

Funded Visioning Activities

Disaster Management · SEES IT · HealthIT · Interactive Tech · Architecture · XLayer · Robotics · Learning Tech · Open Source · Cyber Physical Systems · Global Development · Theoretical CS · Big Data Computing · NetSE · Spatial Computing

From GPS and Virtual Globes to Spatial Computing-2020

About the workshop

This workshop outlines an effort to develop and promote a unified agenda for Spatial Computing research and development across US agencies, industries, and universities. See the original workshop proposal here.

Spatial Computing

Spatial Computing is a set of ideas and technologies that will transform our lives by understanding the physical world, knowing and communicating our relation to places in that world, and navigating through those places.

The transformational potential of Spatial Computing is already evident. From Virtual Globes such as Google Maps and Microsoft Bing Maps to consumer GPS devices, our society has benefitted immensely from spatial technology. We've reached the point where a hiker in Yellowstone, a schoolgirl in DC, a biker in Minneapolis, and a taxi driver in Manhattan know precisely where they are, nearby points of interest, and how to reach their destinations. Large

Logistics

Date: Sept. 10th-11th, 2012
Location: Keck Center
Hotel: Liaison Hotel

Steering Committee

Erwin Gianchandani

Hank Korth

Organizing Committee

Peggy Agouris, George Mason University

Walid Aref, Purdue University

Michael F. Goodchild, University of California - Santa Barbara

Workshop Highlights

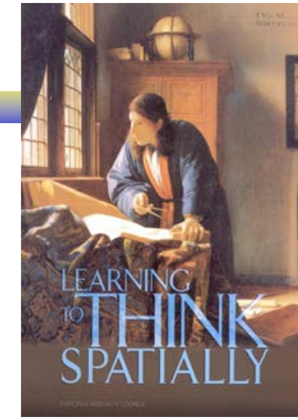
Agenda

- Identify fundamental research questions for individual computing disciplines
- Identify cross-cutting research questions requiring novel, multi-disciplinary solutions

Organizing Committee

- Peggy Agouris, George Mason University
- Walid Aref, Purdue University
- Michael F. Goodchild, University of California - Santa Barbara
- Erik Hoel, Environmental Systems Research Institute (ESRI)
- John Jensen, University of South Carolina
- Craig A. Knoblock, University of Southern California
- Richard Langley, University of New Brunswick
- Ed Mikhail, Purdue University
- Shashi Shekhar, University of Minnesota
- Ouri Wolfson, University of Illinois -
- May Yuan, University of Oklahoma

2012 CCC Workshop: Spatial Computing Visioning

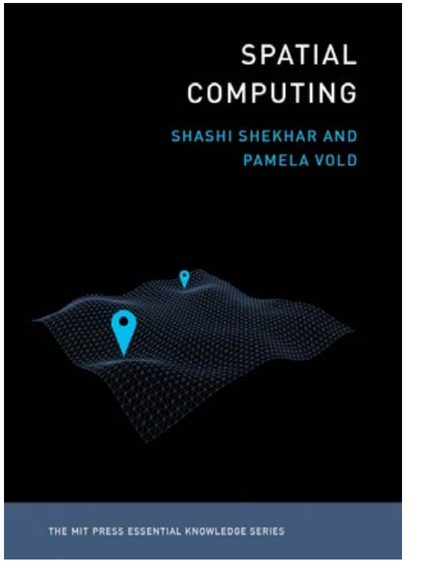


[PDF] Spatial Thinking: A missing building block in STEM education Spatial ...
scienceoflearning.jhu.edu/assets/documents/spatial_thinking_FINAL.pdf
by K Gagnier - Related articles
One critical building block of **success** in **STEM** fields, however, is often overlooked: the ability to think spatially. **Spatial thinking** refers to a set of mental skills that ...

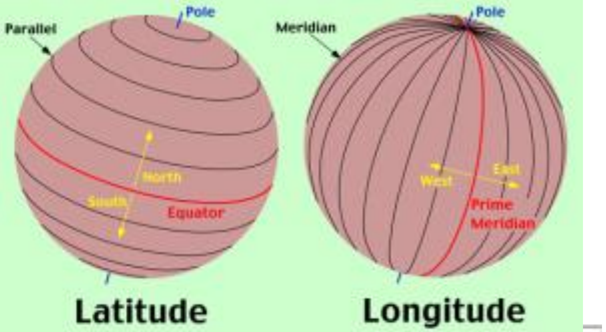
- **Ten Opportunities**
  1. Spatial Abilities Predict STEM Success
  2. Emerging Spatial Big Data
  3. Augmented Reality Systems
  4. Time-Travel in Virtual Globes
  5. Spatial Predictive Analytics
  6. Persistent Environment Hazard Monitoring
  7. Geo-collaborative Systems, Fleets, and Crowds
  8. Localizing Cyber Entities
  9. GPS Deprived Environment
  10. Beyond Geo

Outline

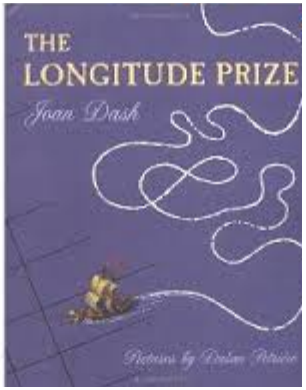
- Introduction
- Broad Interest Examples
  - GPS
    - Outdoors => Indoors
  - Spatial Database Management Systems
  - Location Based Services
  - Spatial Data Science
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions

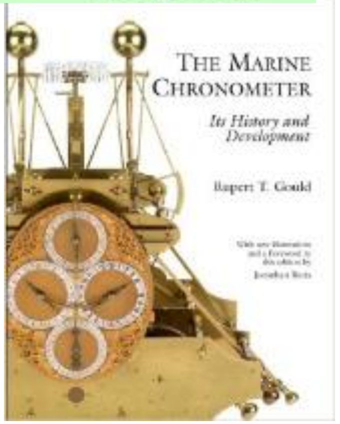


Global Positioning Systems (GPS)

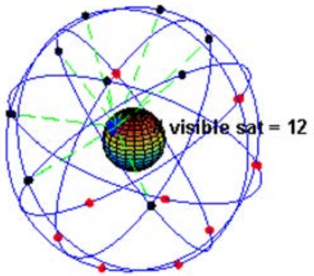


- Positioning ships
  - Latitude f(compass, star positions)
  - Longitude Prize (1714) => marine chronometer
  - accuracy in nautical miles



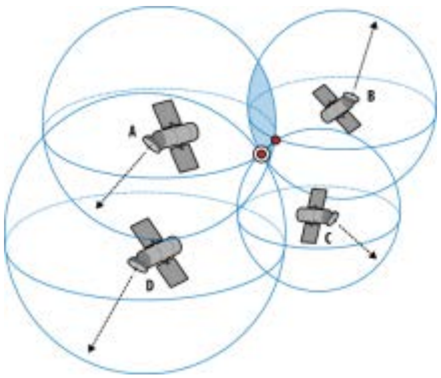


- Global Navigation Satellite Systems
  - Use: Positioning, Clock synchronization
  - Infrastructure: satellites, ground stations, receivers, …

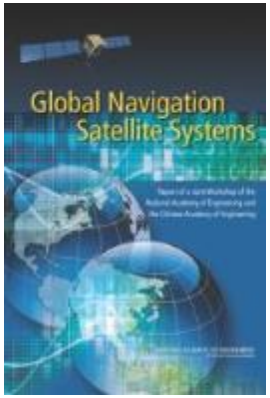


http://en.wikipedia.org/wiki/
Global_Positioning_System

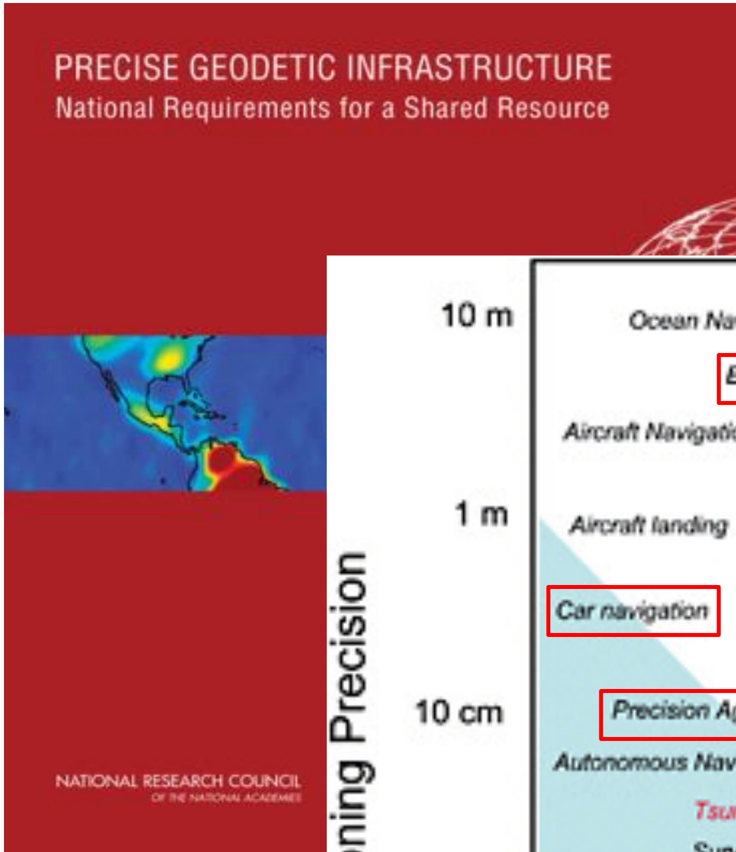
Trilateration

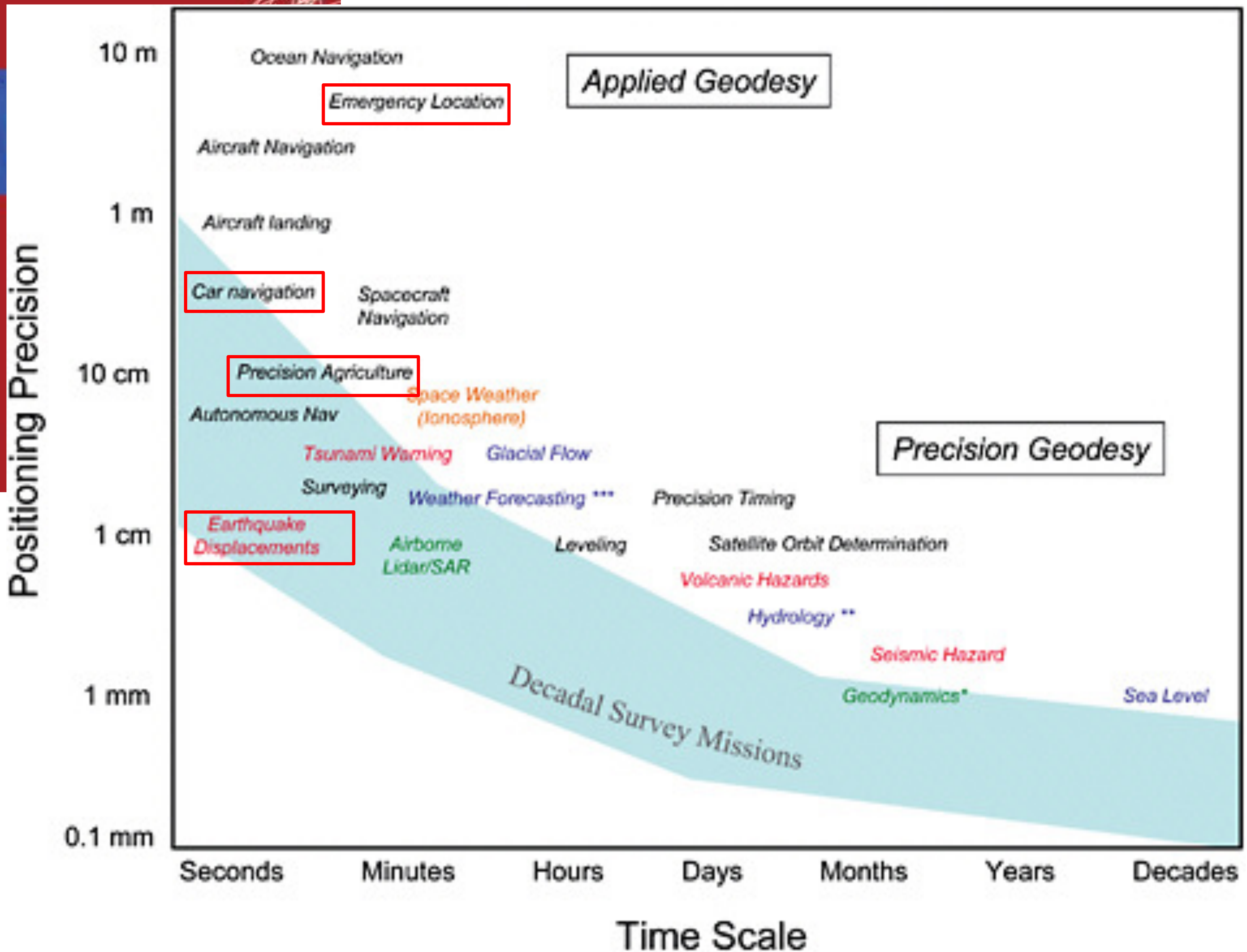


http://answers.oreilly.com/topic/2815
-how-devices-gather-location-
information/



Positioning Precision





Spatial Computing is a Critical Infrastructure Today!

- 2 billion GPS receivers in use, will hit 7 billion by 2022.
- Besides location, it reference time for critical infrastructure
  - Telecommunications industry, Banks, Airlines...
- GPS is the single point of failure for the entire modern economy.
- 50,000 incidents of deliberate (GPS) jamming last two years
  - Against Ubers, Waymo's self-driving cars, delivery drones from Amazon
- Ground based alternatives appearing in S. Korea, USA, …

Bloomberg Businessweek

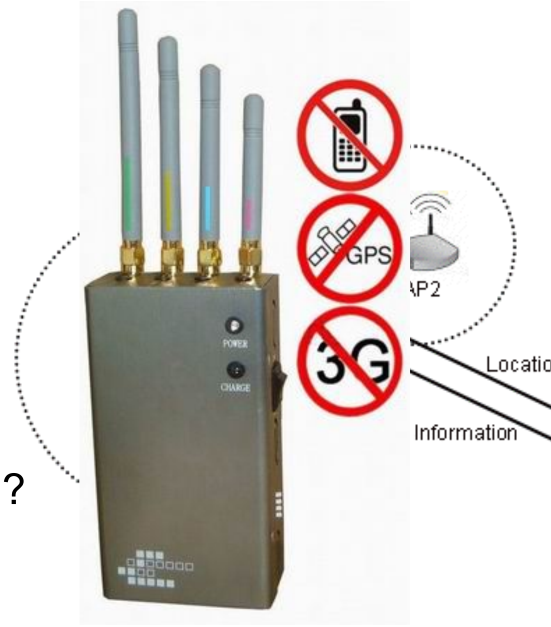
July 25, 2018, 4:00 AM CDT

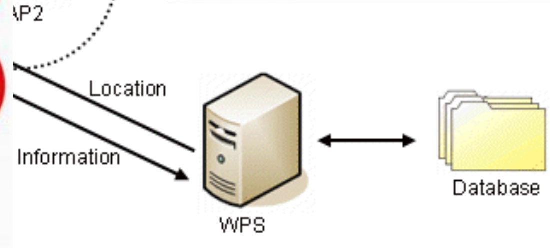
The World Economy Runs on GPS. It Needs a Backup Plan

Source: *https://www.bloomberg.com/news/features/2018-07-25/the-world-economy-runs-on-gps-it-needs-a-backup-plan*

Trends: Localization Without GPS

- GPS works outdoors, but,
  - It can be jammed or spoofed
  - We are indoors 80% of time!
  - Ex. malls, hospitals, airports, …
- Indoor infrastructure
  - Location: Wi-Fi, Blue Tooth, …
  - How to represent indoor for navigation?





Trends: Locate **Cyber** Entities

- Web Server (Internet Node) : Internet Protocol IPv6
- Web-browser: HTML 5
- Voluntary: Checkins on facebook, foursquare, …
- Tweets

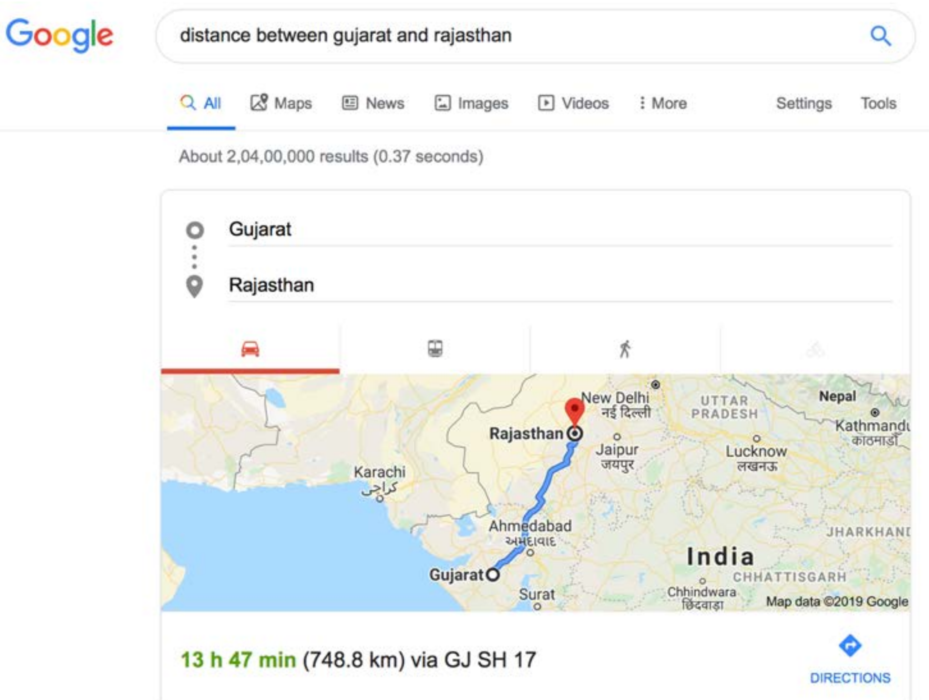
Even before cable news outlets began reporting the tornadoes that ripped through Texas on Tuesday, a **map** of the state **began blinking red** on a screen in the Red Cross' new social media monitoring center, alerting weather watchers that something was happening in the hard-hit area. (AP, April 16th, 2012).

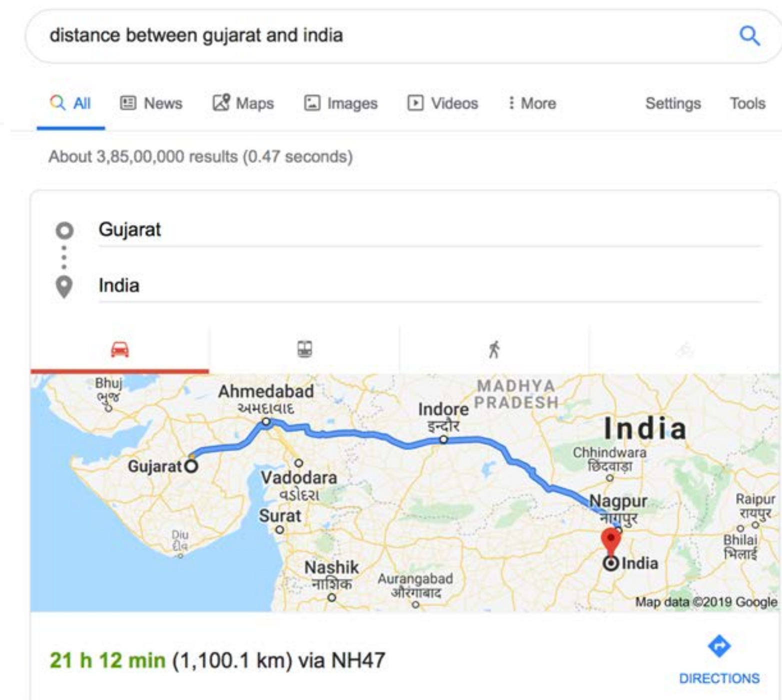
Outline

- Introduction
- Broad Interest Examples
  - GPS
  - Spatial Database Management Systems
    - Point Location => Spatial
    - Scalability => Privacy
  - Location Based Services
  - Spatial Data Science
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions

From (Point) Location to Spatial

- Q? What should Google return for the following questions?
  - Distance between Gujarat and Rajasthan
  - Distance between Gujarat and India



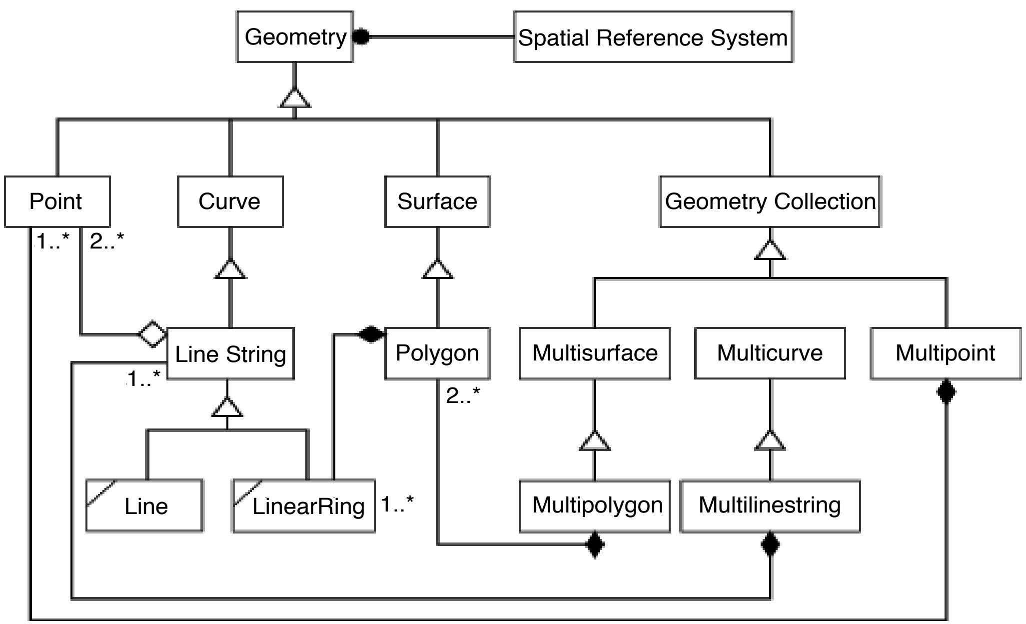


Is GIS just Location?

- Spatial Relationships
  - Ex. Topological, Metric, …
  - OGC Simple Features Standards
- Help feature selection for machine learning & modeling
  - Ex. Distance to key geographic features
  - Ex. Neighbor relationships

Spatial Analysis	Distance
	Buffer
	ConvexHull
	Intersection
	Union
	Difference
	DymmDiff

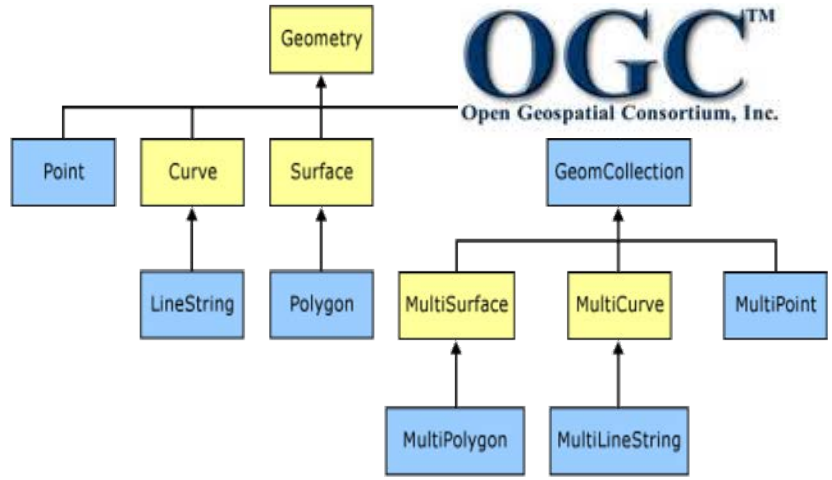
Basic Functions	SpatialReference ()
	Envelop ()
	Export ()
	IsEmpty ()
	IsSimple ()
	Boundary ()
Topological / Set Operators	Equal
	Disjoint
	Intersect
	Touch
	Cross
	Within
	Contains
	Overlap

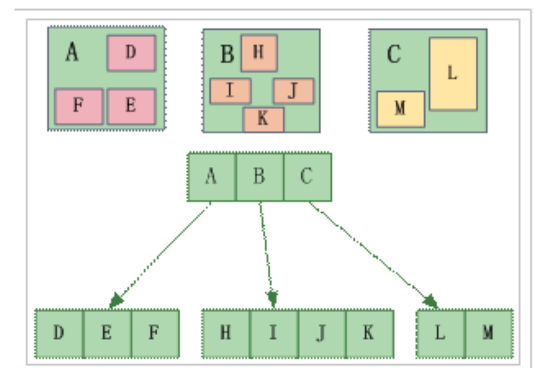


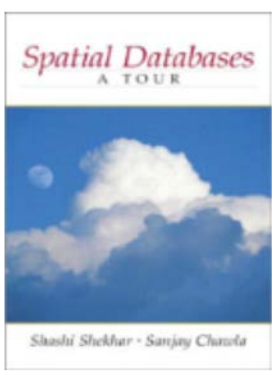
Spatial Big Data Curation

- Meta-data, Schema, DBMS (SQL, Hadoop)
- Challenge: One size does not fit all!

- Ex. Spatial Querying
  - Geo-tag. Checkin, Geo-fence
- Spatial Querying Software
  - OGC Spatial Data Type & Operations
  - Data-structures: B-tree => R-tree
  - Algorithms: Sorting => Geometric
  - Partitioning: random => proximity aware







PostgreSQL

Spatial Hadoop
I can play with your spatial data

Geo-Security & Geo-Privacy

- Operational Security Advice by US Army: Avoid Geo-tags!
  - Q. Why?

Geo-tags can show enemies your location

ArmyTimes Monday Dec 20, 2010

The Army is warning troops to be careful when using Facebook and other popular social networking sites because their geo-tagging features may show where U.S. forces are located in war zones.

Insurgents Used Cell Phone Geotags to Destroy AH-64s in Iraq ...
https://www.defensetech.org › Aircraft
Mar 15, 2012 - From an **Army** press release warning of the dangers of **geotags**: ... location of the **helicopters** inside the compound and conduct a mortar attack, ...

The following was published in Wired Magazine in 2009

“I ran a little experiment. On a sunny Saturday, I spotted a woman in Golden Gate Park taking a photo with a 3G iPhone.

Because iPhones embed geo-data into photos that users upload to Flickr or Picasa, iPhone shots can be automatically placed on a map.

At home I searched the Flickr map, and score—a shot from today. I clicked through to the user’s photostream and determined it was the woman I had seen earlier.

After adjusting the settings so that only her shots appeared on the map, I saw a cluster of images in one location.

Clicking on them revealed photos of an apartment interior—a bedroom, a kitchen, a filthy living room. Now I know where she lives.”

Challenge: Geo-privacy, …

- Emerging personal geo-data
  - Trajectories of smart phones, Google map search, …
- Privacy: Who gets my data? Who do they give it to? What promises do I get?
- Groups: Civil Society, Economic Entities, Public Safety ,Policy Makers

AMBER ALERT

GEOTARGETED
ALERTS AND WARNINGS

Table 4.2: Geo-privacy Policy Conversation Starters

1. Emergencies are different (E-911)
2. Differential geo-privacy can improve saftey (E-911 → PLAN, CMAS)
3. Send apps to data, not vice-versa (e.g., eco-routing)
4. Transparent transactions for location traces for increased consumer confidence
5. Responsible entities for location traces (Credit-bureau/census, HIPPA++ for responsible parties)

Outline

- Introduction
- Broad Interest Examples
  - GPS
  - Spatial Database Management Systems
  - Location Based Services
    - Queries => Persistent Monitoring
  - Spatial Statistics
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions

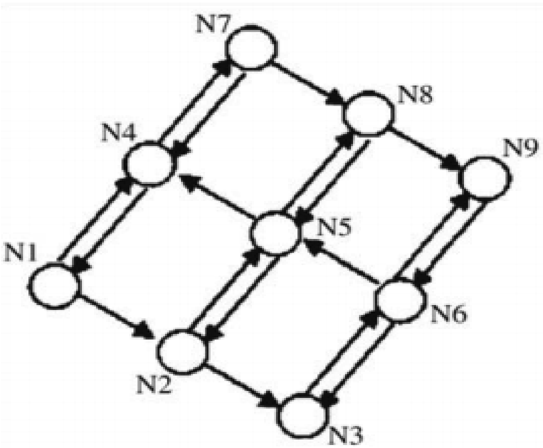
Location Based Services

- <u>Location:</u> Where am I? (street address, <latitude, longitude>)
- <u>Directory:</u>
  - What is around me?
  - Where is the nearest clinic (or ambulance)?
- <u>Routes:</u> What is the shortest path to reach there?

Models: Spatial Graphs & Flow Networks

- Ex.: Roadmaps, Electric grid, Supply chains, …
- Graphs: Nodes, Edges, Routes, …
- Flow networks: Capacity constrain
- Operations:
  - Geo-code, Map-matching, …
  - Connectivity, shortest path, nearest neighbor
  - Logistics: Site selection, Allocation, Max-flow, …

Graph Data for UMN Campus
Courtesy: Bing



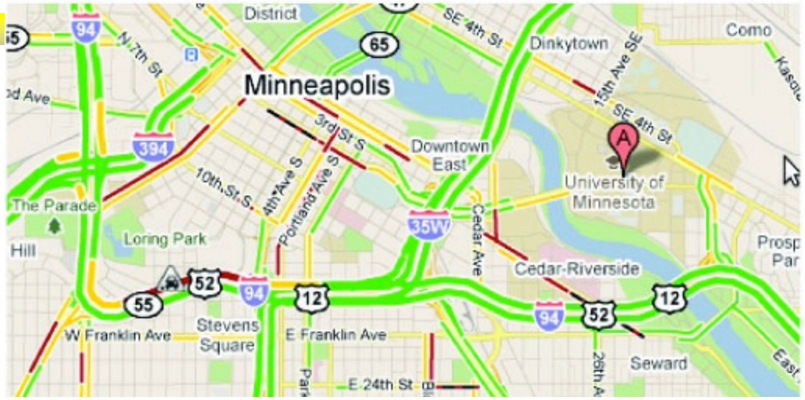
Nodes

NID
N1
N2
N3
N4
N5
N6
N7
N8
N9

Edges

EID	From	To	Speed	Distance
E1	N1	N2	35mph	0.075mi
E2	N1	N4	30mph	0.075mi
E3	N2	N3	35mph	0.078mi
E4	N2	N5	30mph	0.078mi
E5	N3	N6	30mph	0.077mi
E6	N4	N1	30mph	0.075mi
E7	N4	N7	30mph	0.078mi
E8	N5	N2	30mph	0.078mi
…	…	…	…	…

Dynamic Nature of Transportation Network

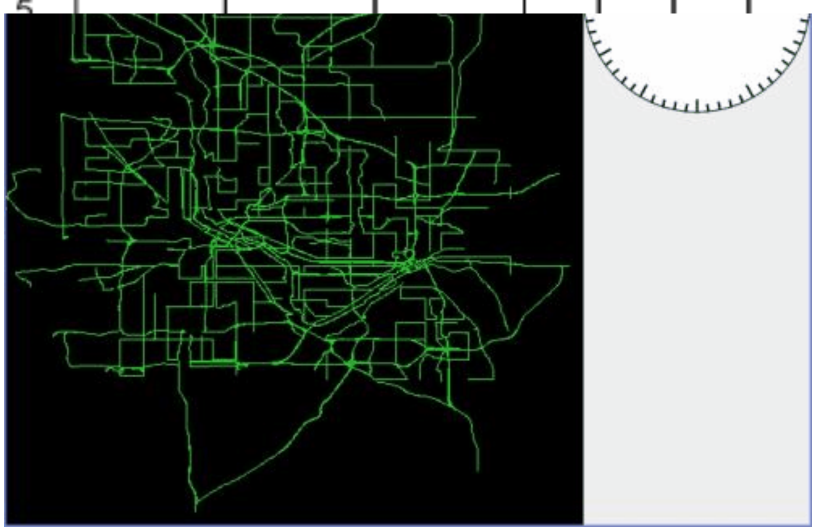


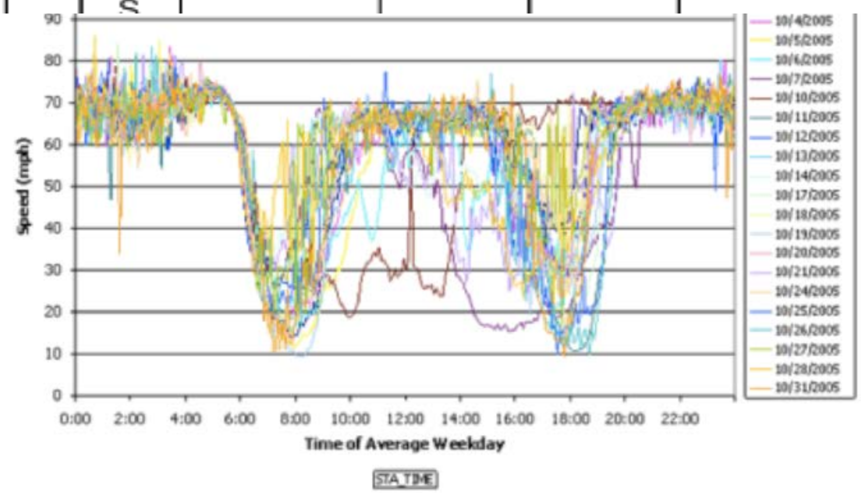
FT_DailyHistoricData

EID	Freeflow Speed	Weekday Speed	Weekend Speed	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	.....	.....	.....	.....	.....	.....	.....	.....	.....	
2	.....	.....	.....	.....	.....	.....	.....	.....	.....	
3	.....	.....	.....	.....	.....	.....	.....	.....	.....	
4	.....	.....	.....	.....	.....	.....	.....	.....	.....	
5										

Historic Daily Speed Profile Table

Speed_0	Speed_1	.....
		.....
		.....





Next Generation Navigation Services

- Eco-Routing
- Best start time
- Road-capacity aware, e.g., evacuation route planning

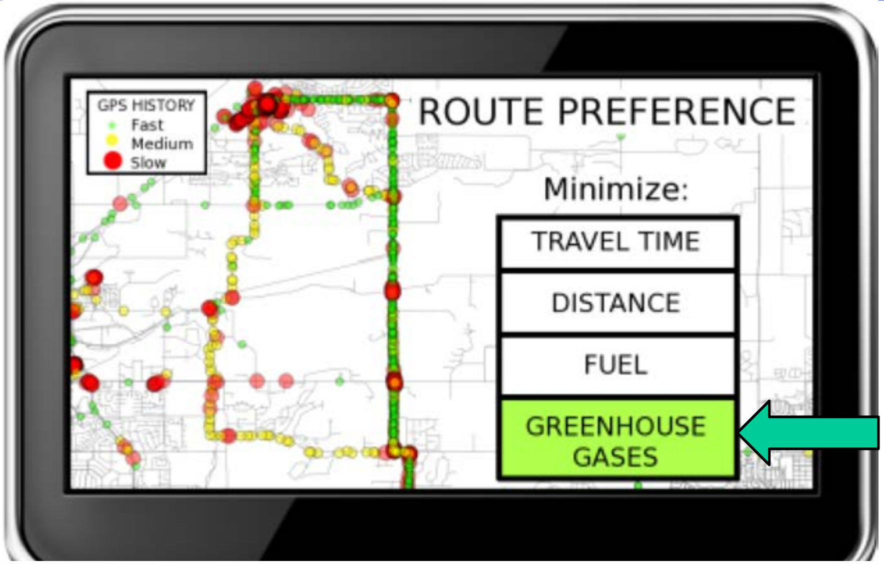
Why UPS trucks (almost) never turn left - CNN.com

www.cnn.com/2017/02/16/world/ups-trucks-no-left-turns/ ▾

Feb 23, 2017 - **Left**-hand **turns** are dangerous and wasteful, data shows. By avoiding them, **UPS** saves 10 million gallons of fuel each year. ... pedestrians than **right** ones, according to data collected by **New**

Next Generation Navigation Services

- Eco-Routing
- Best start time
- Road-capacity aware



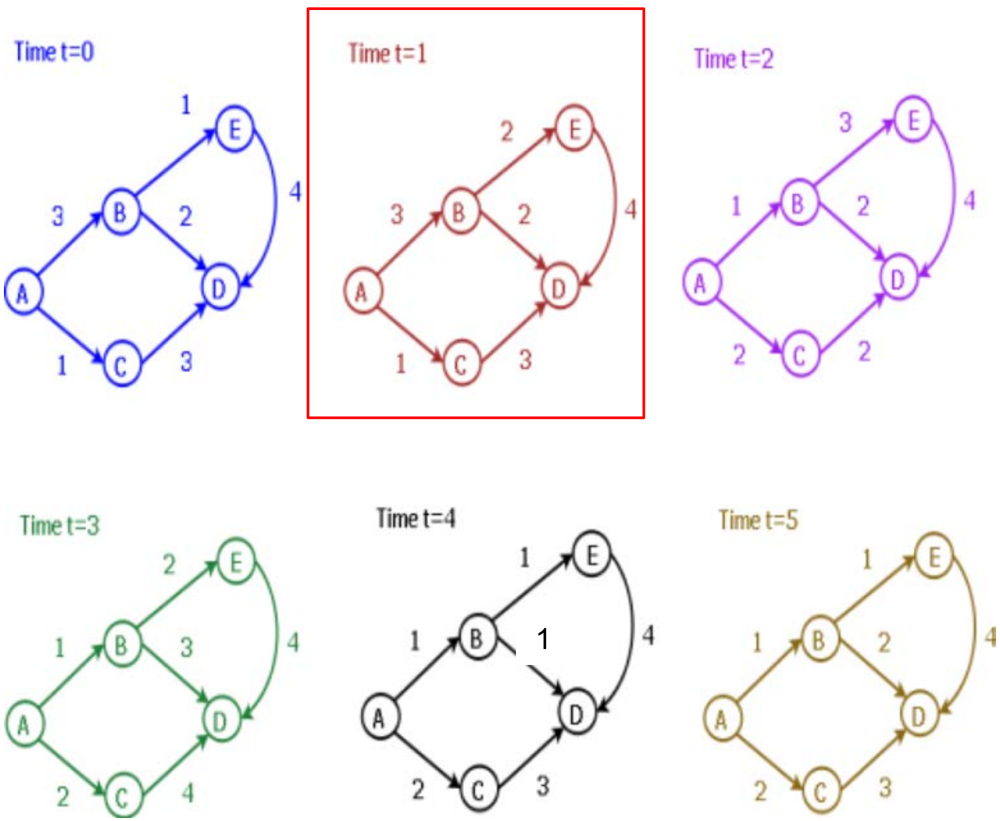
Static	Time-Variant
Which is the shortest travel time path from downtown Minneapolis to airport?	Which is the shortest travel time path from downtown Minneapolis to airport at different times of a work day?
What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis ?	What is the capacity of Twin-Cities freeway network to evacuate downtown Minneapolis at different times in a work day?

Routing Challenges: Lagrangian Frame of Reference

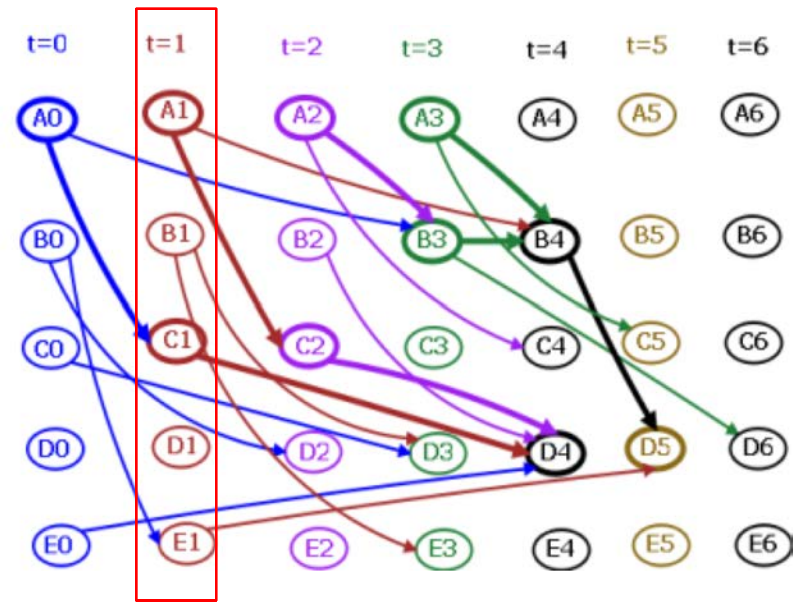
Q? What is the cost of Path <A,C,D> with start-time t=1 ? Is it 3 or 4 ?

Path	T = 0	T = 1	T = 2	T = 3
<A,C,D>	4	3	5	4
<A,B,D>	6	4	4	3

Snapshots of a Graph



Lagrangian Graph



Details: A Critical-Time-Point Approach to All-Start-Time Lagrangian Shortest Paths, IEEE Transactions on Knowledge and Data Engineering, 27(10):2591-2603, Apr. 2015 (A summary in Proc. Intl. Symp. on Spatial and Temporal Databases, Springer LNCS 6849, 2011), (w/ V. Gunturi et al.),

Spatio-temporal Graphs: Computational Challenges

Ranking changes over time

Violates stationary assumption in Dynamic Programming

Time	Preferred Routes
7:30am	Via Hiawatha
8:30am	Via Hiawatha
9:30am	via 35W
10:30am	via 35W

Waits, Non FIFO Behavior

Violate assumption of Dijkstra/A*

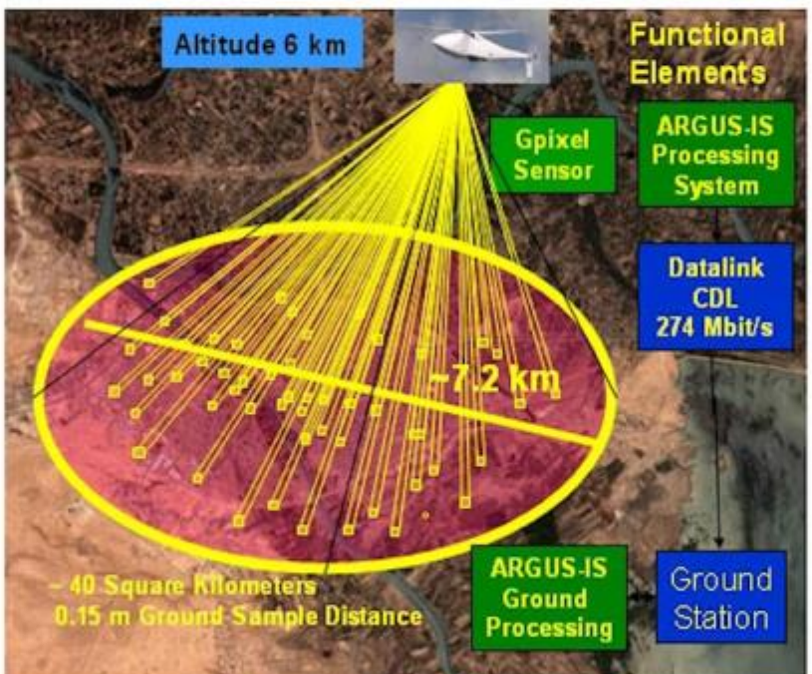
Time	Route	Flight Time
8:30am	via Detroit	6 hrs 31 mins
9:10am	direct flight	2 hrs 51 mins
11:00am	via Memphis	4 hrs 38mins
11:30am	via Atlanta	6 hrs 28 mins
2:30pm	direct flight	2 hrs 51 mins

*Flights between Minneapolis and Austin (TX)

Details: A Critical-Time-Point Approach to All-Start-Time Lagrangian Shortest Paths, IEEE Transactions on Knowledge and Data Engineering, 27(10):2591-2603, Apr. 2015 (A summary in Proc. Intl. Symp. on Spatial and Temporal Databases, Springer LNCS 6849, 2011), (w/ V. Gunturi et al.),

Trends: Persistent Geo-Hazard Monitoring

- Environmental influences on our health & safety
  - air we breathe, water we drink, food we eat
- Large Area Surveillance
  - Passive > Active > Persistent
  - How to economically cover all locations all the time ?
  - Crowd-sourcing, e.g., smartphones, tweets,
  - Wide Area Motion Imagery, UAVs, …

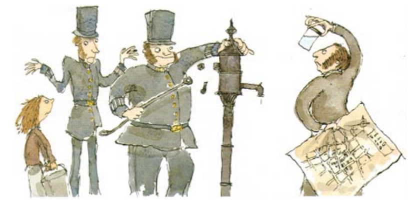


Outline

- Introduction
- Broad Interest Examples
  - GPS
  - Spatial Database Management Systems
  - Location Based Services
  - Spatial Data Science
    - Limitations of Traditional Data Science
    - Novel approaches
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions

Spatial Data Science: A Historical Example

1854: What causes Cholera?

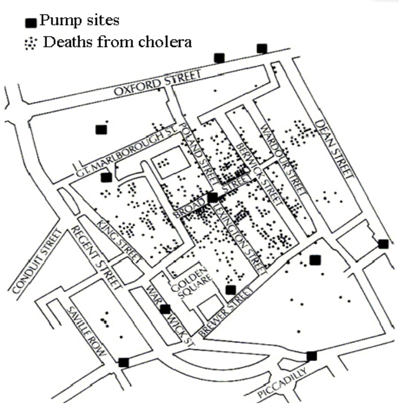


Collect & Curate Data → Discover Patterns, Generate Hypothesis → Test Hypothesis (Experiments) → Develop Theory

? water pump

Remove pump handle

Germ Theory

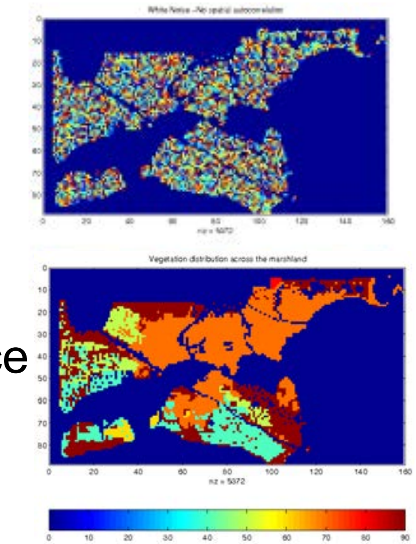


Impact:
sewage system,
drinking water supply
…

Q? What are Choleras of today?
Q? How may Spatial Data Sc. Help?

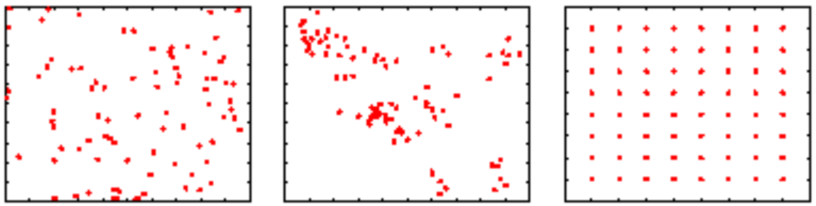
Limitations of Traditional Data Science

- A. High cost of missed or spurious patterns
  - Pr.[Self-driving car sensors fail to detect a red traffic light] > 0
  - Loss of life, stigmatization, economic loss
- B. Gerrymandering risks
  - Spatial partitioning choice may alter results
- C. Spatial data violates assumptions of traditional data science
  - Data samples: independent and identically distributed (i.i.d)
  - Nearby spatial data samples are not independent
  - No two places on Earth are exactly alike!

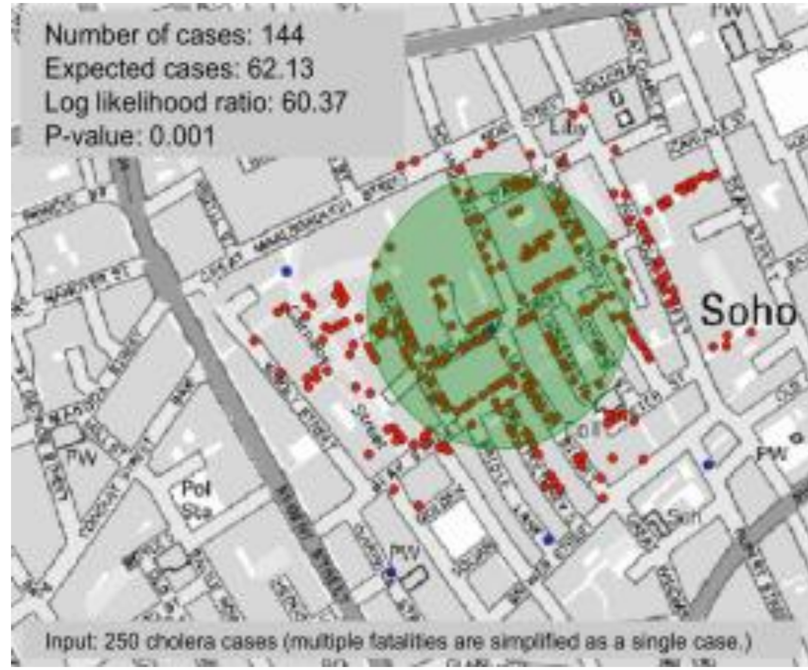


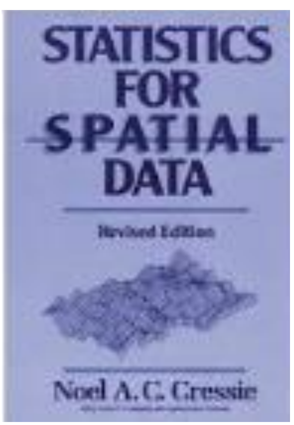
A. Reducing Spurious Patterns

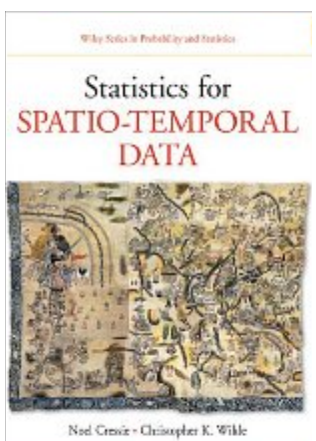
- SatScan (National Cancer Institute)
  - Compare with complete spatial random
  - Monte Carlo simulation



- Spatial Statistics
  - Quantify uncertainty, confidence, …
  - Model Auto-correlation, heterogeneity, …







SaTScan™

Software for the spatial, temporal, and space-time scan statistics

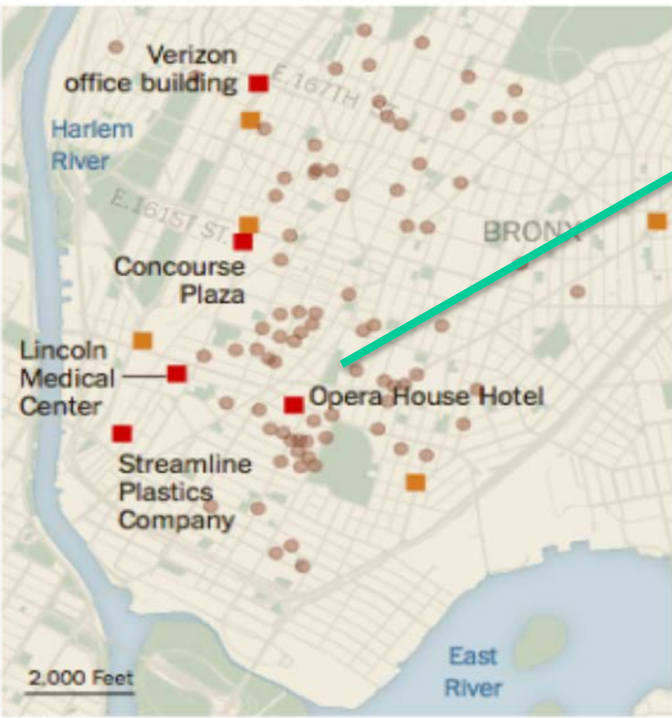
Hotel That Enlivened the Bronx Is Now a 'Hot Spot' for Legionnaires'

By WINNIE HU and NOAH REMNICK AUG. 10, 2015

Contaminated Cooling Towers

Five buildings have been identified as the potential source of the Legionnaires' disease outbreak in the South Bronx.

- Possible sources of Legionnaires' outbreak
- Additional sites found with legionella bacteria
- Locations of people with Legionnaires'

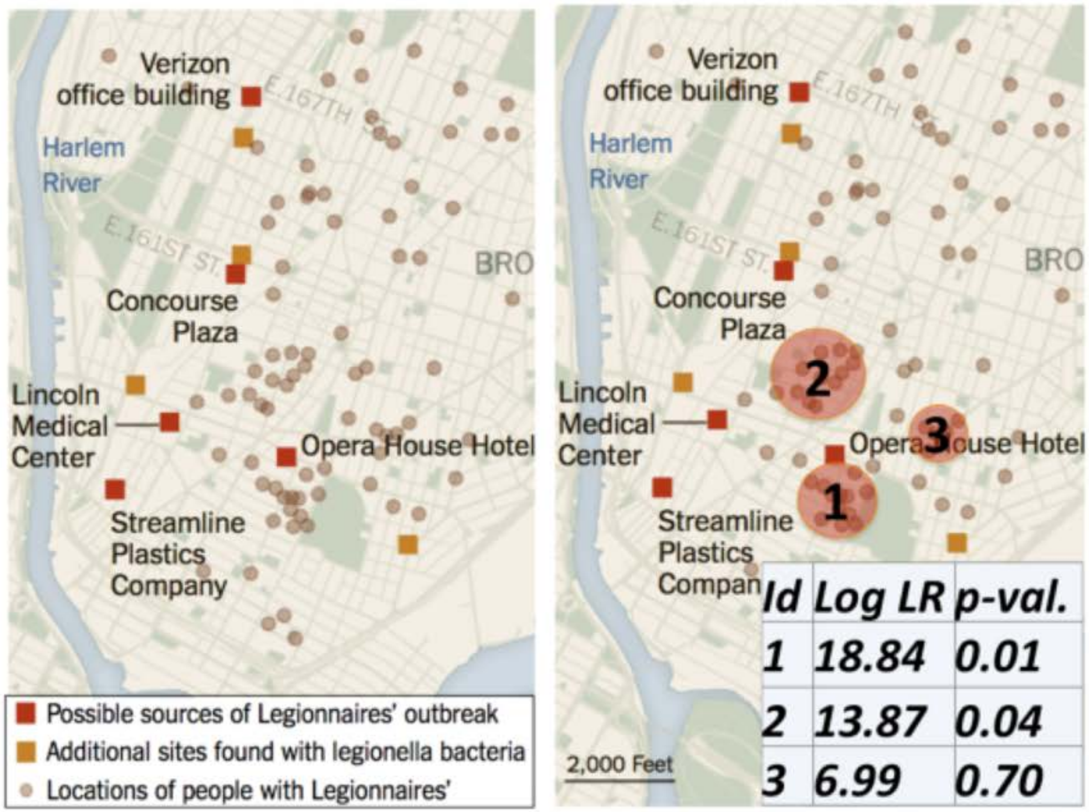


Source: New York Mayor's Office

By The New York Times

The Opera House Hotel is at the center of the outbreak. Edwin J. Torres for The New York Times

Legionnaires' Disease Outbreak in New York



Id	Log LR	p-val.
1	18.84	0.01
2	13.87	0.04
3	6.99	0.70

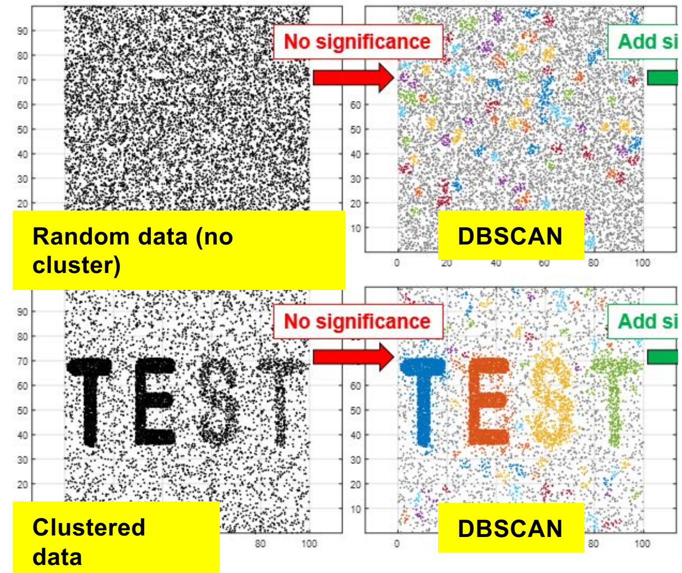
(a) Legionnaire's in New York (2015) (b) Output of SaTScan

Details: Ring-Shaped Hotspot Detection. IEEE Trans. On Data & Knowledge Eng., 28(120:3367-3381 , Dec. 2016 (: A Summary in Proc. IEEE ICDM 2014) (w/ E. Eftelioglu et al.)

Significant Hotspot (Arbitrary Shape)

Problem definition

- **Inputs:** A set of points; DBSCAN parameters; Test statistic; Significance level
- **Output:** Significant clusters
- **Objective:** Computational efficiency

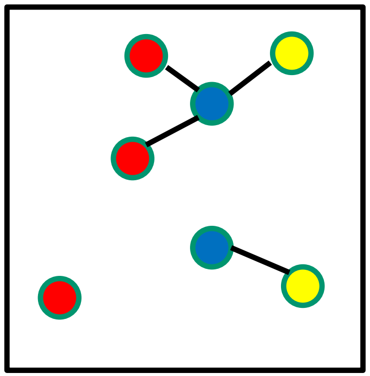


- **Contributions**
  - Significance modeling in DBSCAN
  - A fast dual-convergence algorithm

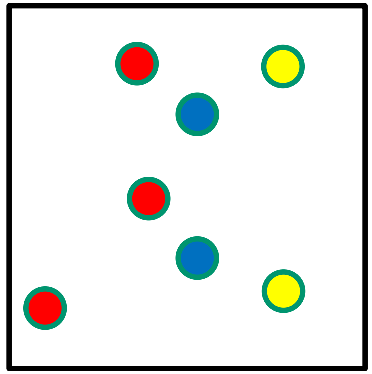
- **Trends**
  - DBSCAN cannot avoid chance patterns
  - SaTScan cannot detect arbitrary shapes

Details: Significant DBSCAN towards Statistically Robust Clustering, Y. Xie and S. Shekhar.Proc. 16th International Symposium on Spatial and Temporal Databases (SSTD '19), 2019, ACM **(Best Paper Award)**

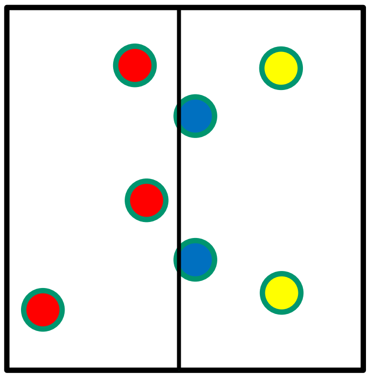
B. Neighbor Graph Reduces Gerrymandering Risks



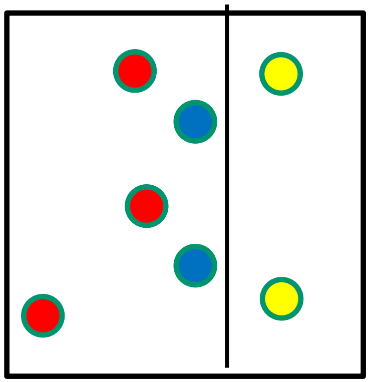
(d) Neighbor graph



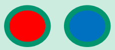
(a) a map



(b) Partition A



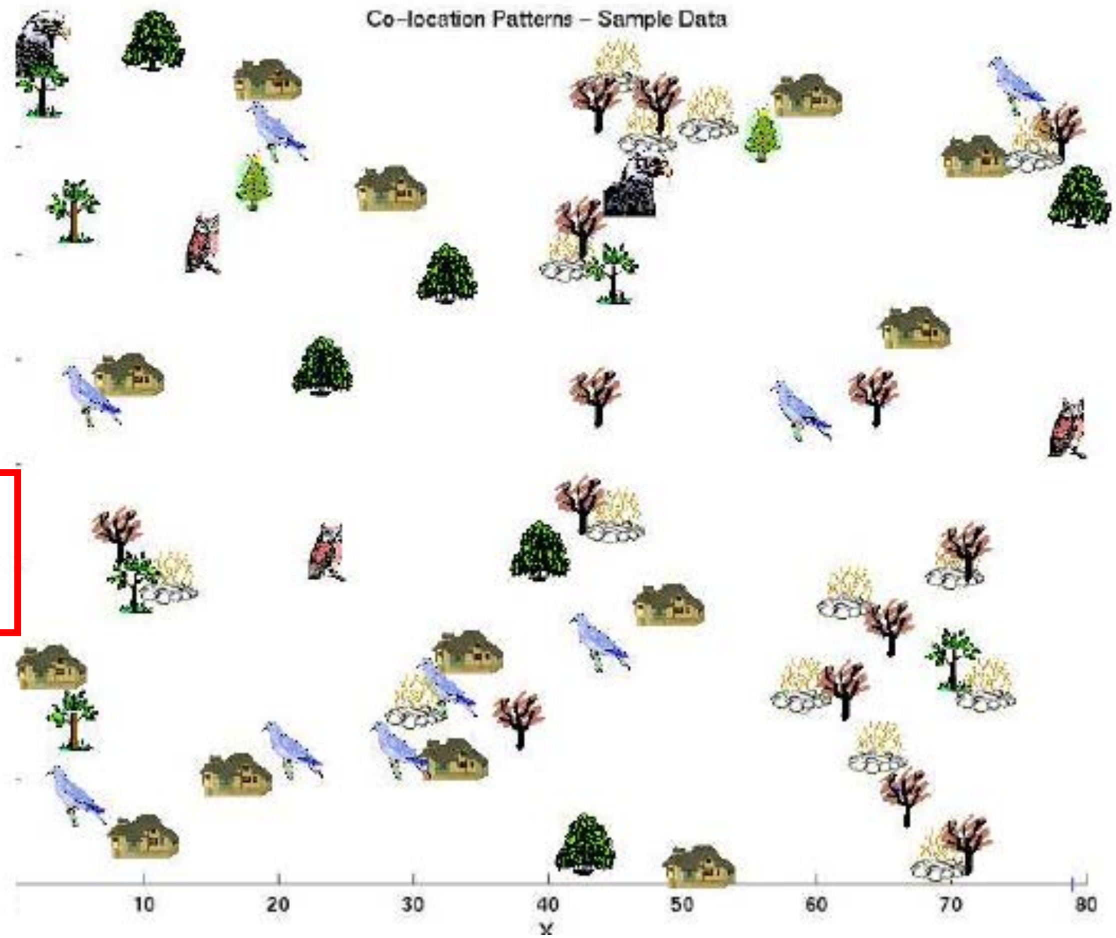
(c) Partition C

Participation Index	Ripley's Cross-K	Pattern	Pearson Correlation	Pearson Correlation
0.5	0.33		(-) 0.9	1
1	1		1	(-) 0.9

Spatial Computing Research Group

Co-locations/Co-occurrence

- Given: A collection of different types of spatial events
- Find: Co-located subsets of event types



Details: Discovering colocation patterns from spatial data sets: a general approach, (w/ H. Yan et al.), IEEE Transactions on Knowledge and Data Engineering, 16(12), Dec. 2004.

Colocation Mining

Participation ratio $pr(f_i, c)$ of feature f_i in colocation $c = \{f_1, f_2, \ldots, f_k\}$:

fraction of instances of f_i with feature $\{f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_k\}$ nearby

Participation index $PI(c) = \min\{ pr(f_i, c) \}$

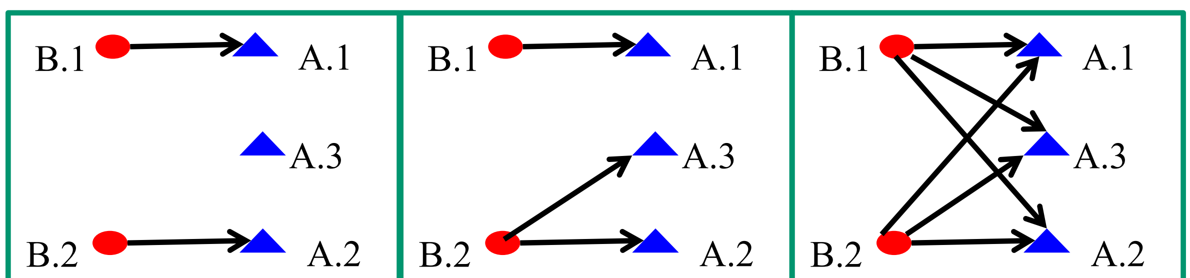
Properties:

(1) Computational: Non-monotonically decreasing like support measure

Allows scaling up to big data via pruning

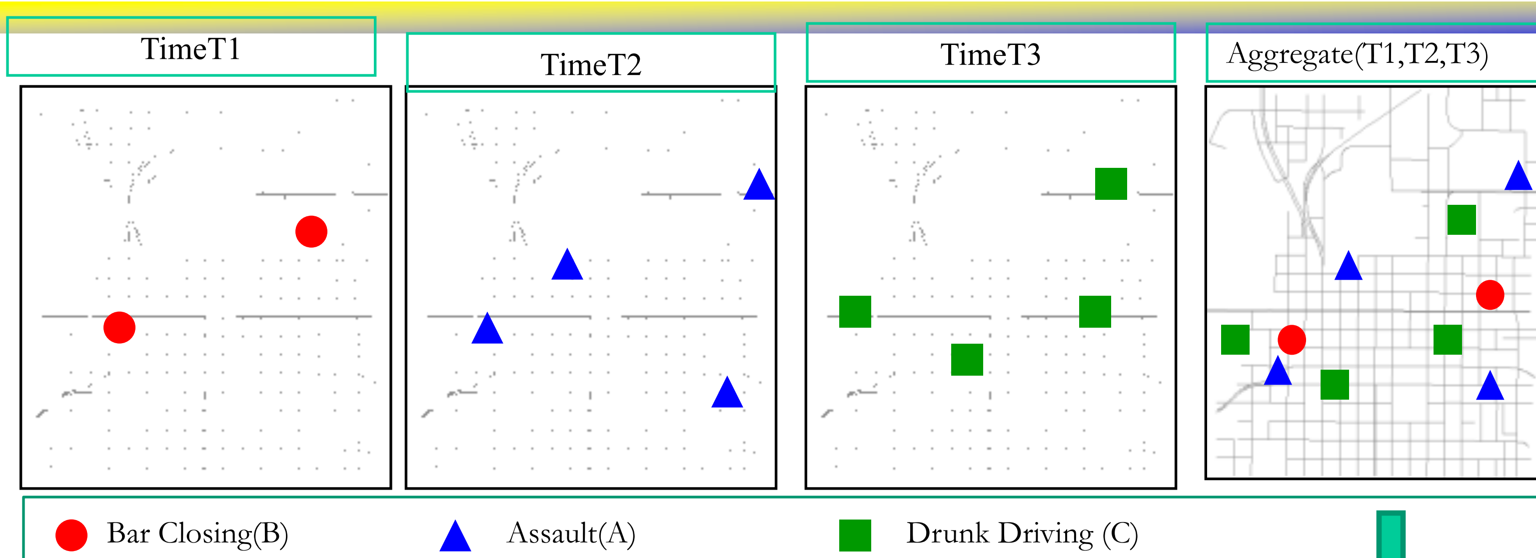
(2) Statistical: Upper bound on Cross-K function

- Comparison with Ripley's K-function (Spatial Statistics)

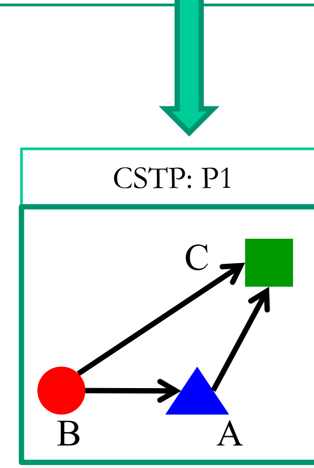


K-function (B , A)	2/6 = 0.33	3/6 = 0.5	6/6 = 1
PI (B , A)	2/3 = 0.66	1	1

Cascading spatio-temporal pattern (CSTP)

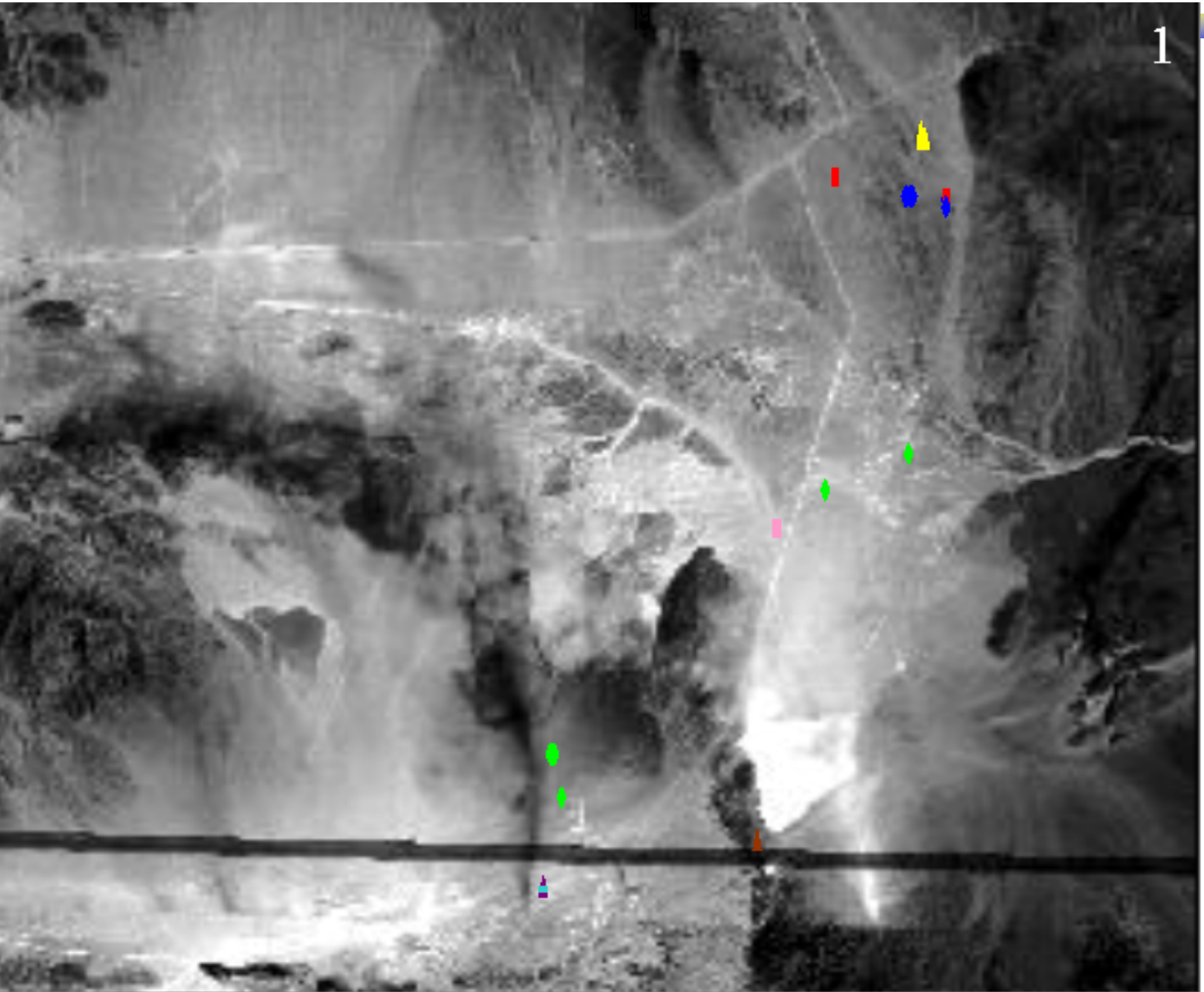


- *Input:* Urban Activity Reports
- *Output: CSTP*
    - *Partially ordered* subsets of ST event types.
    - Located together in space.
    - Occur in *stages* over time.
- Applications: Public Health, Public Safety, …



Details: Cascading Spatio-Temporal Pattern Discovery, (w/ P. Mohan et al.), IEEE Transactions on Knowledge and Data Engineering, 24(11), Nov. 2012.

MDCOP Motivating Example : Input



- Manpack stinger (2 Objects)

- M1A1_tank (3 Objects)

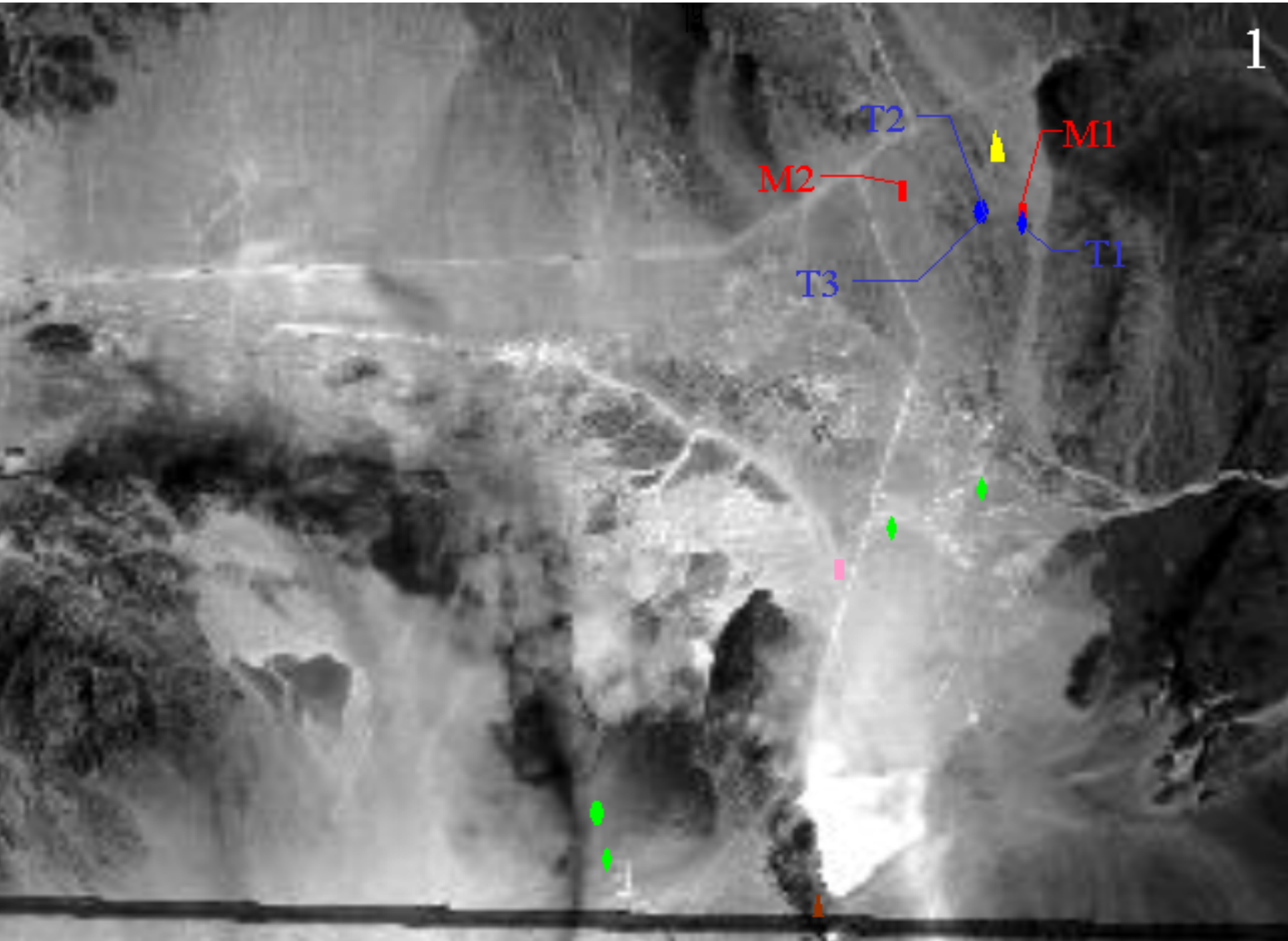
- M2_IFV (3 Objects)

- Field_Marker (6 Objects)
- T80_tank (2 Objects)

- BRDM_AT5 (enemy) (1 Object)

- BMP1 (1 Object)

MDCOP Motivating Example : Output



Details: Mixed-Drove Spatiotemporal Co-Occurrence Pattern Mining,, (w/ M. Celik et al.)
IEEE Transactions on Knowledge and Data Engineering, 20(10), Oct. 2008.

- Manpack stinger (2 Objects)

- M1A1_tank (3 Objects)

- M2_IFV (3 Objects)

- Field_Marker (6 Objects)
- T80_tank (2 Objects)

- BRDM_AT5 (enemy) (1 Object)

- BMP1 (1 Object)

C1. Modeling Auto-correlation in Prediction Models

- Traditional Models
  - Linear Regression (e.g., Logit), Bayes Classifier, Neural Networks, Decision Trees
- Semi-Spatial : auto-correlation in regularizer
- Spatial Models
  - W = neighbor matrix (row-normalized)
  - Spatial autoregressive model (SAR)
  - Markov random field (MRF) based Bayesian Classifier

$$\varepsilon = \|y - X\beta\|^2 + \|y - y_{neighbor}\|^2$$

Traditional	Spatial
$y = X\beta + \varepsilon$	$y = \rho W y + X \beta + \varepsilon$
$\Pr(C_i \mid X) = \dfrac{\Pr(X \mid C_i) \Pr(C_i)}{\Pr(X)}$	$\Pr(c_i \mid X, C_N) = \dfrac{\Pr(C_i) \Pr(X, C_N \mid c_i)}{\Pr(X, C_N)}$

Ex.: Spatial Auto-Regression Parameter Estimation

ρ : the spatial auto-regression (auto-correlation) parameter

$\mathbf{W}$: n-by-n neighborhood matrix over spatial framework

Name	***Model***
Classical Linear Regression	$\boxed{\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$
Spatial Auto-Regression	$\boxed{\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$

- **Maximum Likelihood Estimation** $$\ln(L) = \boxed{\ln\left|\mathbf{I} - \rho\mathbf{W}\right|} - \frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - SSE$$

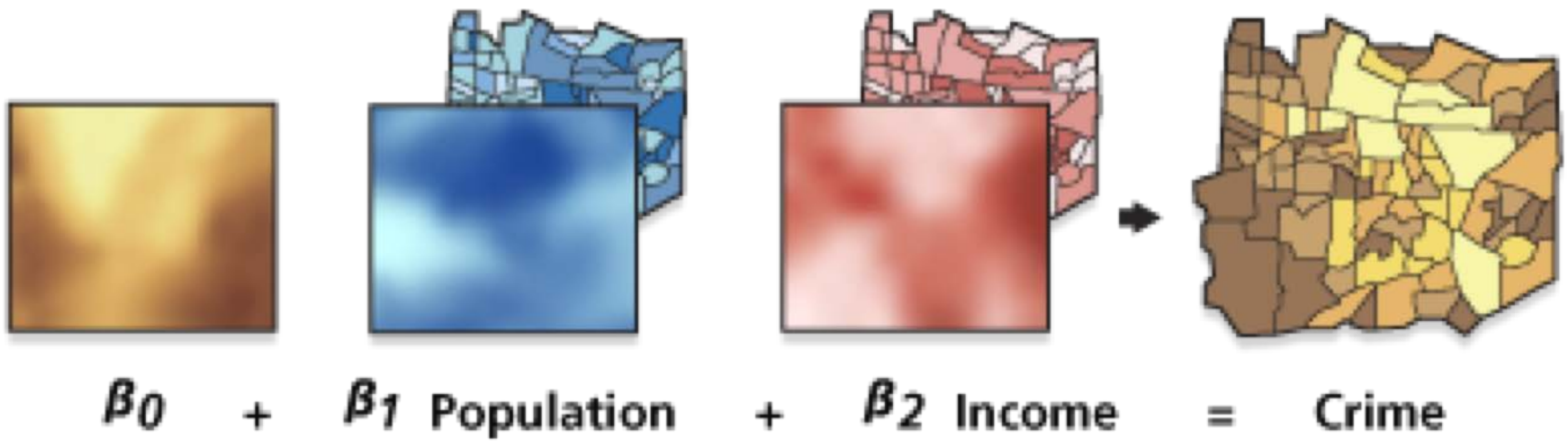
- Computing determinant of large matrix is a hard (open) problem!
  - size(W) is quadratic in number of locations/pixels.
  - Typical raster image has Millions of pixels
  - W is sparse but not banded.

A parallel formulation of the spatial auto-regression model for mining large geo-spatial datasets, SIAM Intl. Workshop on High Perf. and Distr. Data Mining, 2004. (w/ B. Kazar)

C2. Modeling Spatial Heterogeneity: GWR

- Geographically Weighted Regression (GWR)
  - Goal: Model spatially varying relationships
  - Example: $y = X\beta^{'} + \varepsilon^{'}$

    Where $\beta^{'}$ and $\varepsilon^{'}$ are location dependent



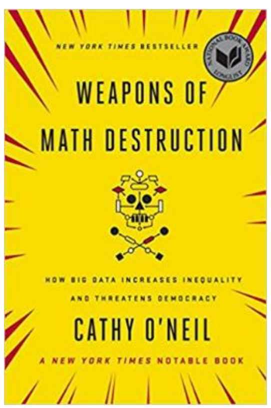
Source: resources.arcgis.com

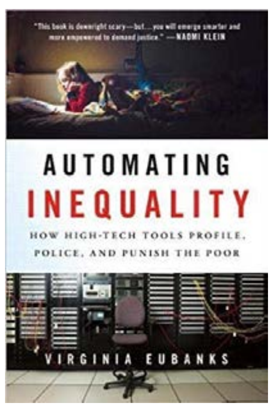
Quiz

- Which are addressed in Convolutional Neural Networks (CNN) ?
  - Statistical significance to reduce chance Patterns
  - Spatial Auto-correlation
  - Spatial Heterogeneity

Trends: Civil Society Concerns

- Was Tesla "self-driving" claim fatal? : https://www.youtube.com/watch?v=o02H2xGIecc
- Are automated face recognition software fair?
- NSF DCL 19-016: Fairness, Ethics, Accountability, and Transparency: Enabling Breakthrough Research to Expand Inclusivity in CISE Research
- Books:
  - Weapons of Math Destruction, Cathy O'Neil, 2016 (2019 Euler Book Prize, TED talk)
  - Automating Inequality, V. Eubanks, 2018.





CARS

Tesla drivers' autopilot fails are going viral on YouTube

Brad Reed @bwreedbgr
October 22nd, 2015 at 8:30 AM
Share Tweet

OK, everyone, repeat after me: Tesla's new Autopilot feature is not a substitute for driving. That nonetheless hasn't stopped many Tesla drivers from relying too heavily on Autopilot and posting videos of scary experiences of it on YouTube. Here are some of the best/worst examples we could find.

Self-Driving Cars Still Can't Handle Snow, Rain, or Heavy Weather

By Joel Hruska on October 30, 2018 at 4:53 pm | 87 Comments

Outline

- Introduction
- Broad Interest Examples
  - GPS
  - Spatial Database Management Systems
  - Location Based Services
  - Spatial Data Science
  - Virtual Globes & Remote Sensing
    - Quilt => Time-travel & Depth
  - Geographic Information Systems
- Conclusions

These High-Tech Sensors May Be the Key to Autonomous Cars

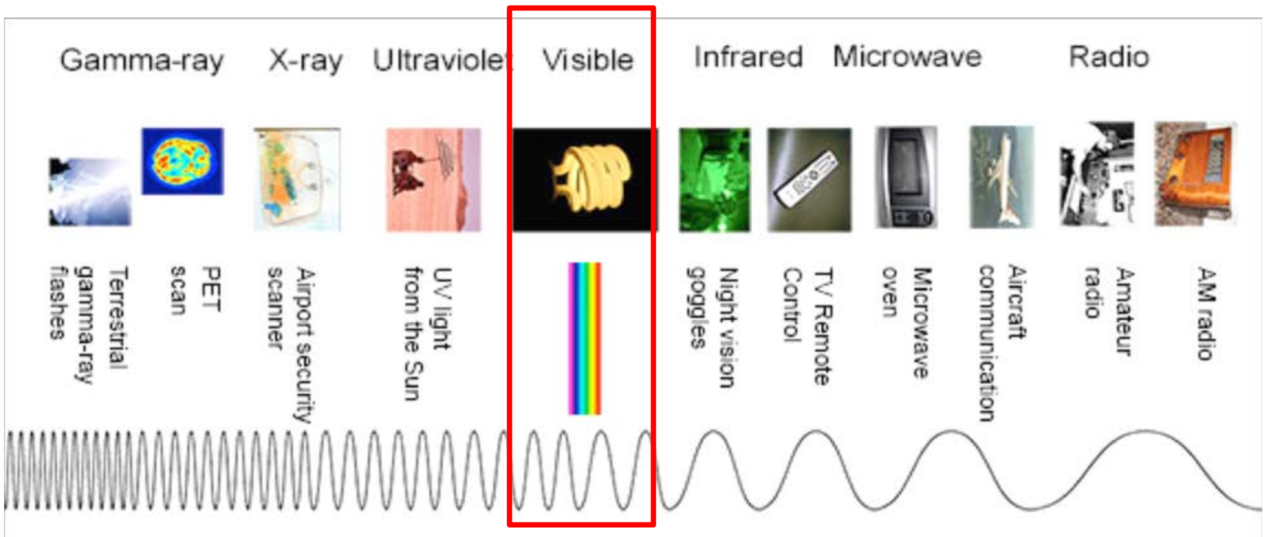
The New York Times Sept. 26, 2019

A downpour in Las Vegas at the CES technology show a year and a half ago may prove to have been a watershed moment in the race to develop autonomous cars. While other companies promoting their experimental self-driving vehicles had to keep them parked in the rain, one company, AdaSky, demonstrated how its sensors could see people hundreds of feet ahead even in a downpour, and even when they were wearing white and standing against a white background.

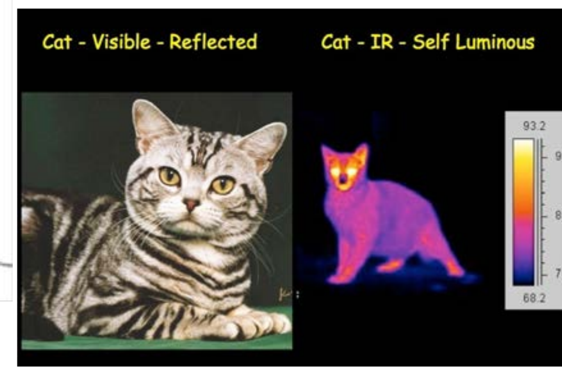
"Thermal imaging is the best sensor at detecting people, day or night," Chris Posch of FLIR Systems said.

ElectroMagnetic (EM) radiation

- Emitted by all objects above absolute zero (0° Kelvin(K), −273°C)



Source: imagine.gsfc.nasa.gov

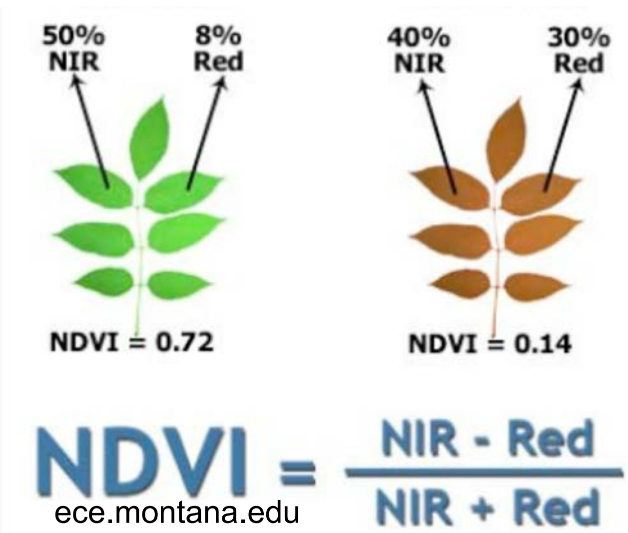


Source: directthermography.co.uk

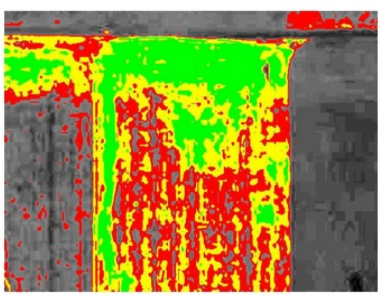
Spectral Indices

- Index = a summary across multiple bands to predict a feature, e.g., vegetation, water, …
- Example using near-infra-red (NIR) and red (RED)
  - Ratio vegetation index (RVI) or Simple ratio index (SRI) = NIR/RED
  - Normalized Difference Vegetation Index (NDVI) = (NIR - RED) / (NIR + RED)
  - Physical interpretation: energy absorption, photosynthetic capacity

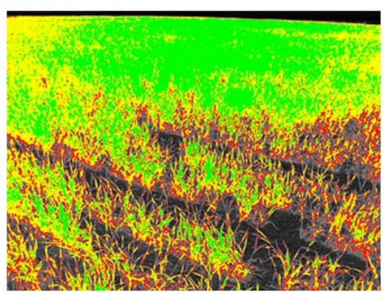
- Ex.: NDVI for Healthy vegetation (NIR = 50%, RED 8%)
  - Stressed/sparse vegetation (NIR = 40%, RED = 30%)
  - Q? How may a farmer use NDVI to monitor crops?



ece.montana.edu



Agribotix UAV-collected VI imagery recognized that the wheat density observed from the road was not indicative of the whole field. The red areas will likely produce less wheat.



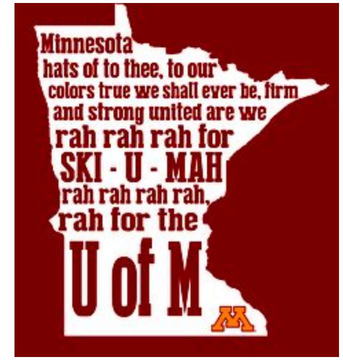
We walked to this location to ground-truth the aerial images and found much sparser rows in the red areas shown in the image at left.

Bestdroneforthejob.com

Spatial Computing Research Group

Virtual Globes & Volunteered Geo-Information

- Virtual Globes: Geo distribution, patterns
  - 1995: UMN Map Server
  - 1998: Al Gore's Digital Earth Speech
  - 1999: Microsoft Terra-server
  - 2004: Keyhole (Google Earth) : Fly-through

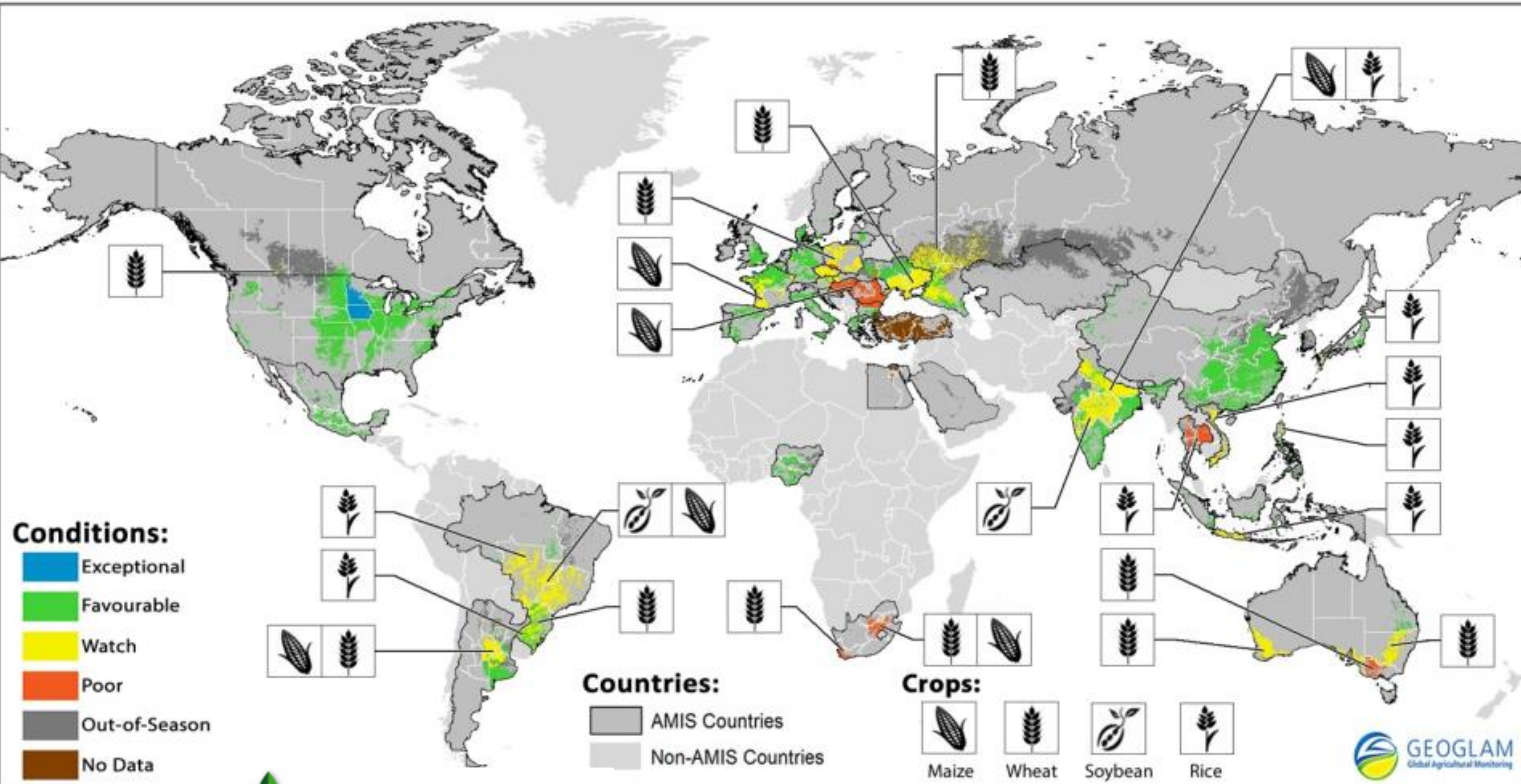


The Enduring Vision of a Digital Earth

Speech by Al Gore, Jan. 31, 1998

- Volunteered Geo-Information
  - Allow citizens to make maps & report
  - 2009 Haiti Post-Earthquake Maps
  - Road maps, Traffic maps, …

Remote Sensing – Agriculture Monitoring



AMIS Agricultural Market Information System

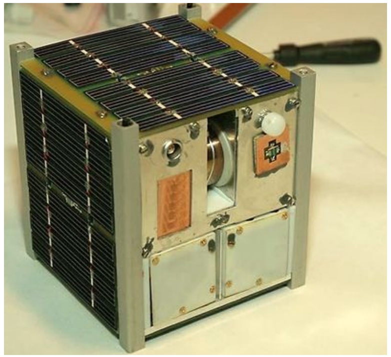
GEOGLAM Global Agricultural Monitoring

Opportunities: Time-Travel and Depth in Virtual Globes

- Virtual globes are (quilt) snapshots
- How to add time?
  - Ex. NASA NEX, Google Earth Engine,
  - Ex. Google Timelapse: 260,000 CPU cor hours for global 29-frame video

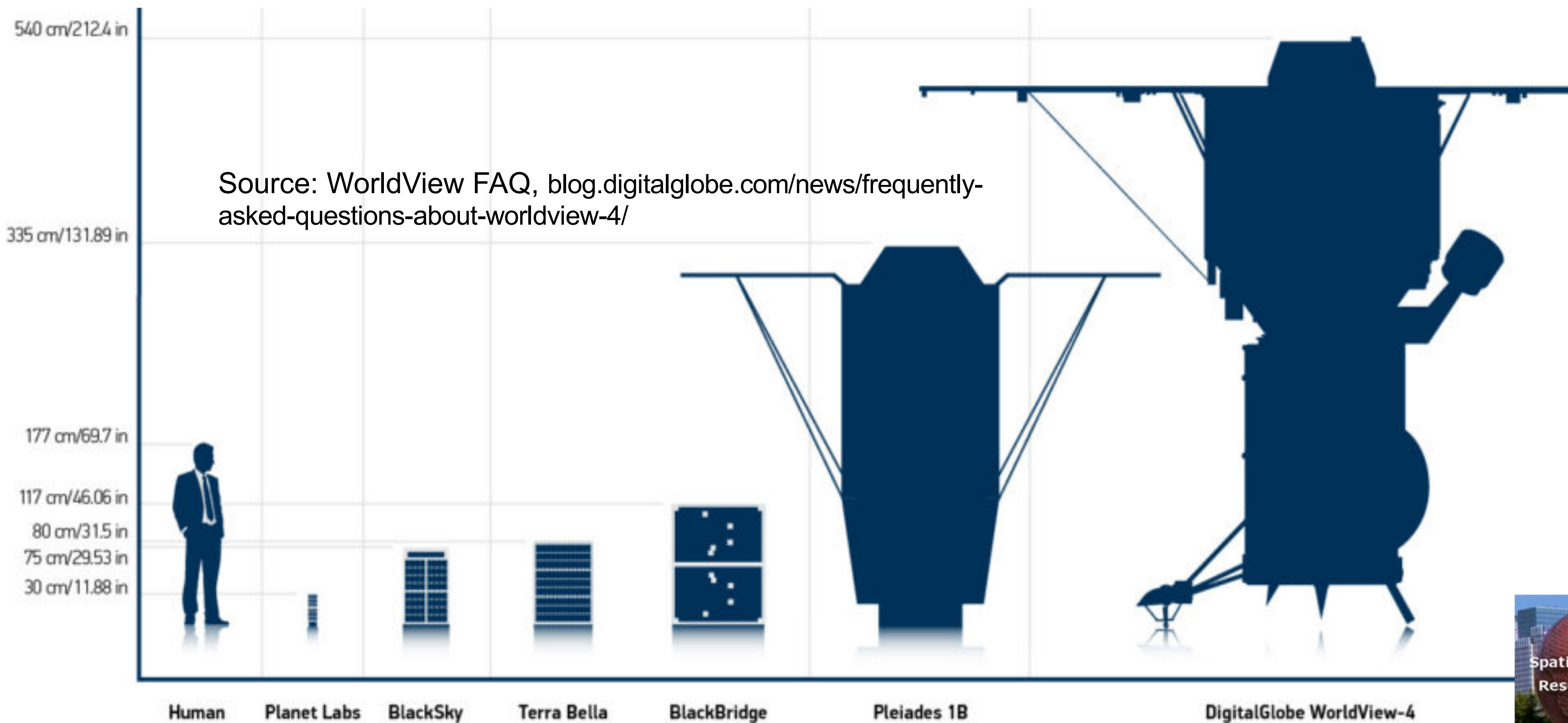
googleblog.blogspot.com/2013/05/a-picture-of-earth-through-time.html

- Spatio-temporal Resolution
  - Planet Labs. : daily 1m scan (visual bands)
  - USDA VegScape / CropScape
- Small Satellites
  - CubeSat (10cm x 10cm x 11.35cm)



Large Constellations of Small Satellites

- Hi-frequency (e.g., daily or hourly) time-series of imagery of entire earth
  - Monitor illegal fishing, forest fires, crops (2017 DARPA Geospatial Cloud Analytics)
- Small Satellites: video (5-minutes): https://geospatialstream.com/sciencecasts-nasa-embraces-small-satellites/
- Large Constellations
  - 2017: Planet Labs: 100 satellites: daily scan of Earth at 1m resolution in visible band



Cheap (or free) satellite data on cloud computers

- 2008: USGS gave away 35-year LandSat satellite imagery archive
  - Analog of public availability of GPS signal in late 1980s
- 2017: Many cloud-based Virtual collaboration environment
  - Explosion in machine learning on satelliite imagery to map crops, water, buildings, roads, …

	Google Earth Engines	NEX	AWS Earth
Elevation, Landsat, LOCA, MODIS, NAIP	x	x	x
NOAA	x		x
AVHRR, FIA, GIMMM, GlobCover, NARR, TRIMM, Sentinel-1	x	x	
IARPA, GDELT, MOGREPS, OpenStreetMap, Sentinel-2, SpaceNet (building/road labels for ML)			x
CHIRPS, GeoScience Australia, GSMap, NASS, Oxford Map, PSDI, WHRC, WorldClim, WorldPop, WWF,	x		
BCCA, FLUXNET		x	

Spatial Computing Research Group

From Earth Observation to Geo-Dashboards

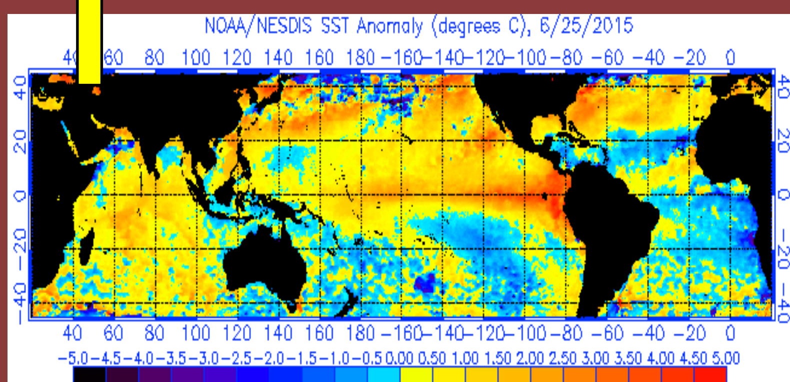
Aral Sea Shrinkage (1978-2014)
Due to Cotton Farms

Alerts

Geo Dashboard

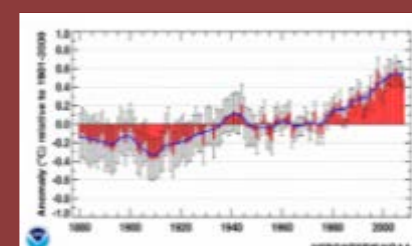
State

Trends

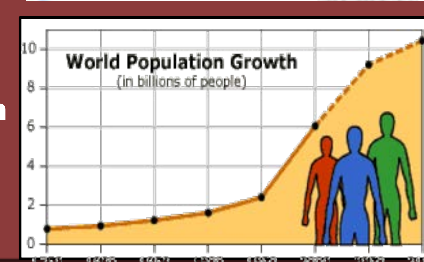


Sea-Surface Temperature Anomaly

Global Temperature



Global Population



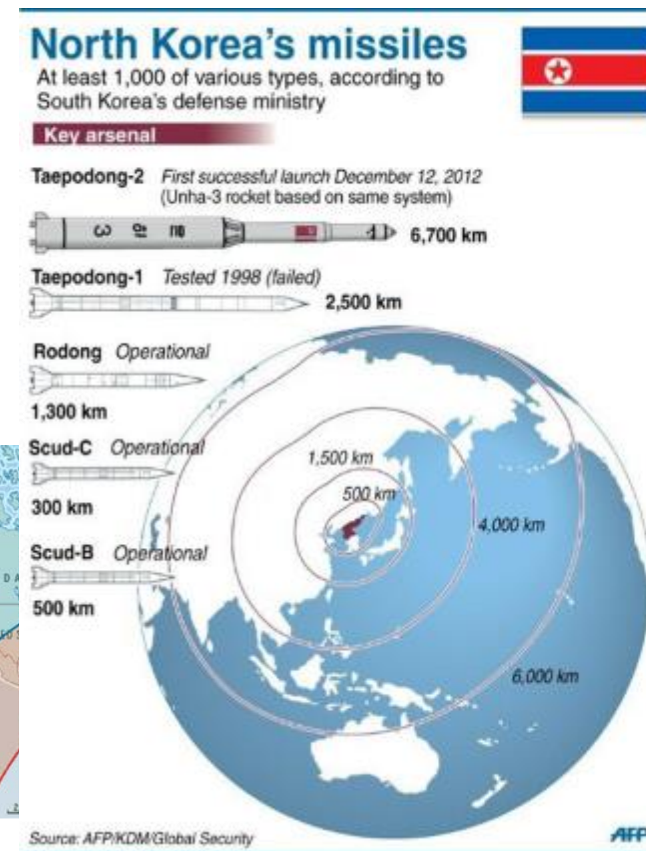
Outline

- Introduction
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Data Science
  - Spatial Database Management Systems
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
    - Geo => Beyond Geo
- Conclusions

Geographic Information Systems & Geodesy

- **GIS**: An umbrella system to
  - capture, store, manipulate, analyze, manage, and present diverse geo-data.
  - SDBMS, LBS, Spatial Statistics, …
  - Cartography, Map Projections, Terrain, etc.

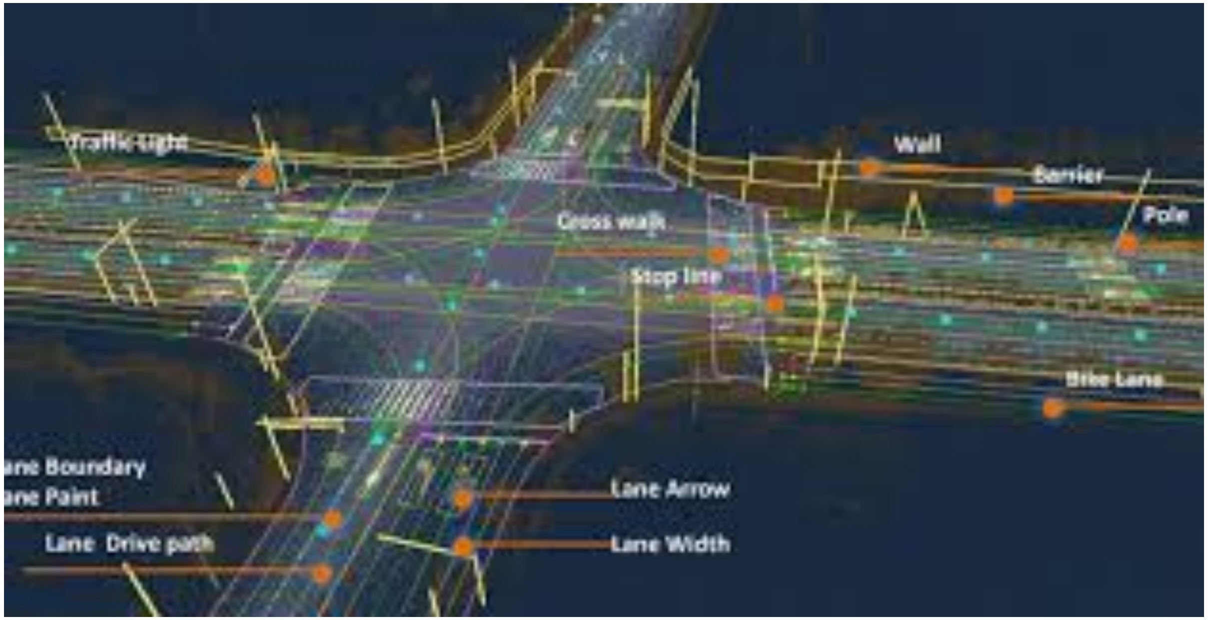
- Map Projections
  - Which countries in North Korea missile range?
  - Spherical coordinates vs. its planar projections

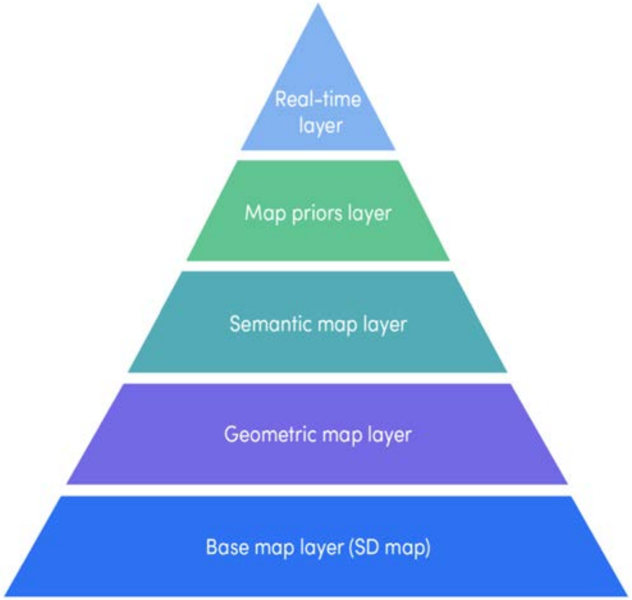


Original The Economist Correction

Trend: 3 Dimensions, e.g., High Definition (HD) Roadmaps

- Trucks and bridges (https://www.youtube.com/watch?v=USu8vT_tfdw)
- Self- Driving Cars, HD-Roadmaps (Cm accuracy)
  - Base Map: road centerlines
  - (3D) Geometry: overpass, walls, curbs, slope, …
  - Semantic: known traffic lights, stop sign, …
  - Map priors: known object types (e.g., people, cars, …
  - Real-time: Camera, LiDAR, …

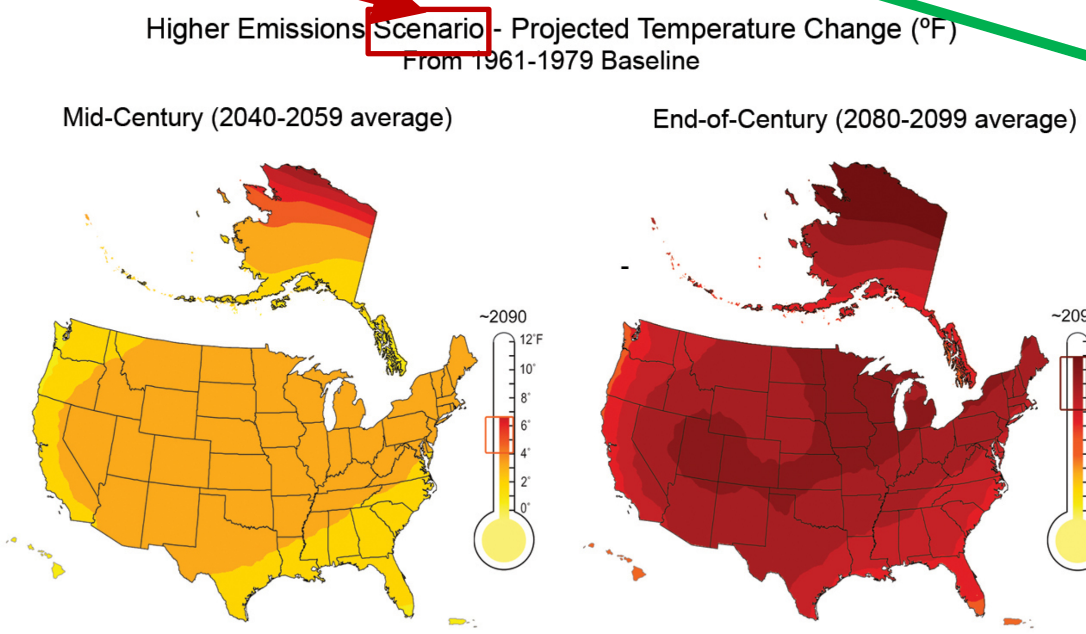


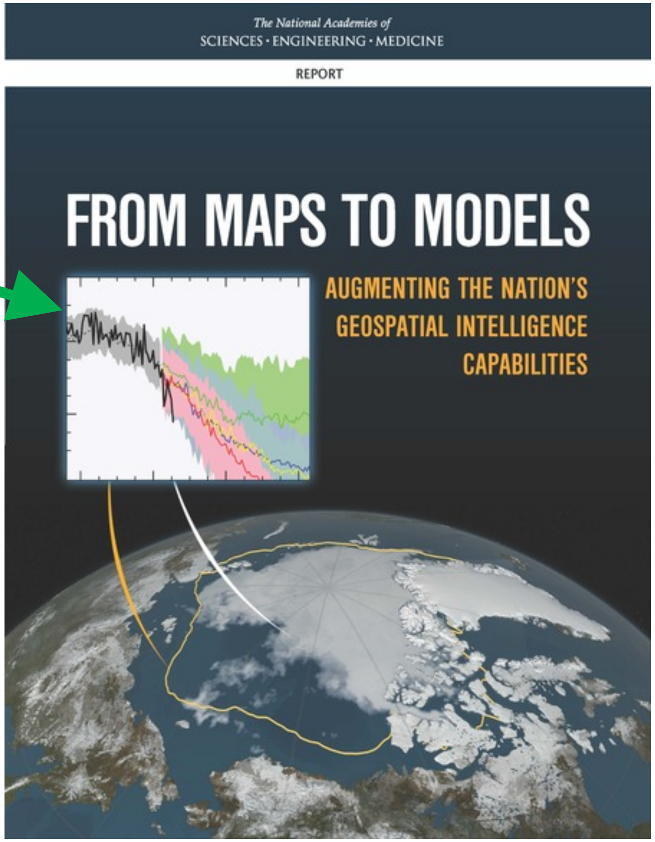


https://m.futurecar.com/3346/BMW-China-and-Beijing-based-Mapping-Company-NavInfo-to-Develop-HD-Maps-for-Autonomous-Driving

Trends: From Maps to Models

- Ex. Climate Projections
- Uncertainty Quantification
  - Scenarios, Confidence Bands





Facilitate Collaboration & Interaction

- Example: Collaborative Geodesign

- Goal: Improving water quality under limited budget

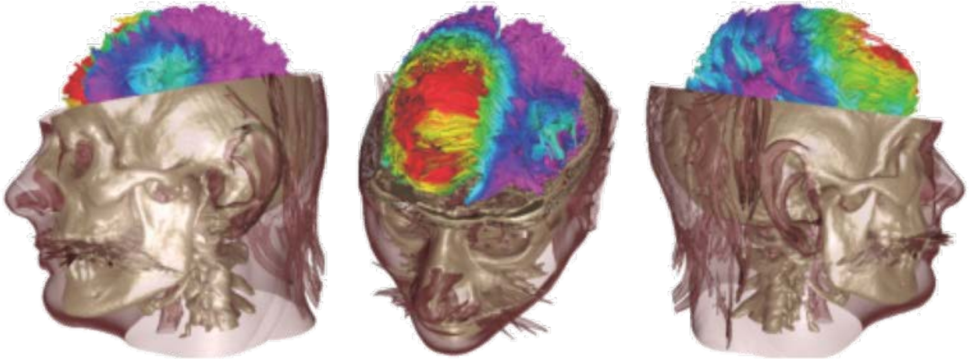
- Features
  - **Collaboration** to resolve conflicts
  - **Interactive** land allocation
  - Real-time visualization and feedback
  - Iterate till convergence

Opportunities: Beyond Geographic Space

- Spaces other than Earth
  - Challenge: reference frame?
- Ex. Human body
  - What is Reference frame ?
    - Adjust to changes in body
    - For MRIs, X-rays, etc.
  - What map projections?
  - Define path costs and routes to reach a brain tumor ?

Outer Space	Moon, Mars, Venus, Sun, Exoplanets, Stars, Galaxies
Geographic	Terrain, Transportation, Ocean, Mining
Indoors	Inside Buildings, Malls, Airports, Stadiums, Hospitals
Human Body	Arteries/Veins, Brain, Neuromapping, Genome Mapping
Micro / Nano	Silicon Wafers, Materials Science

http://convergence.ucsb.edu/issue/14



Oliver, Dev, and Daniel J. Steinberger. "From geography to medicine: exploring innerspace via spatial and temporal databases." Advances in Spatial and Temporal Databases. Springer Berlin Heidelberg, 2011. 467-470.

Outline

- Introduction
- GPS
- Spatial Database Management Systems
- Location Based Services
- Spatial Data Science
- Virtual Globes & Remote Sensing
- Geographic Information Systems
- Conclusions

Summary

The World Economy Runs on GPS.

- Spatial Data are ubiquitous & important
- Spatial Computing has transformed our society
  - –It is only a beginning!
  - –It promises an astonishing array of opportunities in coming
- Current Data Science Tools are inadequate
  - – Gerrymandering, Spatial Auto-correlation, …
- **Ask:** Data Science Degrees should include
  - –Spatial Data Science Methods…

A UCGIS Call to Action:

Bringing the Geospatial Perspective to Data Science Degrees and Curricula

University Consortium for
GEOGRAPHIC INFORMATION SCIENCE

Summer 2018

References :Surveys, Overviews

- Spatial Computing, MIT Press Essential Knowledge Series, 2020. (S. Shekhar, P. Vold)
- Spatial Computing (html , short video , tweet), Communications of the ACM, 59(1):72-81, January, 2016.
- Transdisciplinary Foundations of Geospatial Data Science (html , pdf), ISPRS Intl. Jr. of Geo-Informatics, 6(12):395-429, 2017. (doi:10.3390/ijgi6120395)
- Spatiotemporal Data Mining: A Computational Perspective , ISPRS Intl. Jr. on Geo-Information, 4(4):2306-2338, 2015 (DOI: 10.3390/ijgi4042306).
- Identifying patterns in spatial information: a survey of methods (pdf), Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3):193-214, May/June 2011. (DOI: 10.1002/widm.25).
- Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data, IEEE Transactions on Knowledge and Dat Mining, 29(10):2318-2331, June 2017. (DOI: 10.1109/TKDE.2017.2720168).
- Parallel Processing over Spatial-Temporal Datasets from Geo, Bio, Climate and Social Science Communities: A Research Roadmap. IEEE BigData Congress 2017: 232-250.
- Spatial Databases: Accomplishments and Research Needs, IEEE Transactions on Knowledge and Data Engineering, 11(1):45-55, 1999.

References: Details

Colocations	• Discovering colocation patterns from spatial data sets: a general approach, *IEEE Trans. on Know. and Data Eng.*, 16(12), 2004 (w/ Y. Huang et al.). • A join-less approach for mining spatial colocation patterns, IEEE Trans. on Know. and Data Eng.,18 (10), 2006. (w/ J. Yoo). • Cascading Spatio-Temporal Pattern Discovery. IEEE Trans. Knowl. Data Eng. 24(11): 1977-1992, 2012 (w/ P. Mohan et al.).
Spatial Outliers	• Detecting graph-based spatial outliers: algorithms and applications (a summary of results), Proc.: ACM Intl. Conf. on Knowledge Discovery & Data Mining, 2001 (with Q. Lu et al.) • A unified approach to detecting spatial outliers, Springer GeoInformatica, 7 (2), 2003. (w/ C. Lu, et al.) • Discovering Flow Anomalies: A SWEET Approach, IEEE Intl. Conf. on Data Mining, 2008 (w/ J. Kang).
Hot Spots	• Discovering personally meaningful places: An interactive clustering approach, ACM Trans. on Info. Systems (TOIS) 25 (3), 2007. (with C. Zhou et al.) • A K-Main Routes Approach to Spatial Network Activity Summarization, IEEE Trans on Know. & Data Eng., 26(6), 2014. (with D. Oliver et al.) • Significant Linear Hotspot Discovery, IEEE Trans. Big Data 3(2): 140-153, 2017, (w/ X.Tang et al.)
Location Prediction	• Spatial contextual classification and prediction models for mining geospatial data, IEEE Transactions on Multimedia, 4 (2), 2002. (with P. Schrater et al.) • Focal-Test-Based Spatial Decision Tree Learning. IEEE Trans. Knowl. Data Eng. 27(6): 1547-1559, 2015 (summary in Proc. IEEE Intl. Conf. on Data Mining, 2013) (w/ Z. Jiang et al.).
Change Detection	• Spatiotemporal change footprint pattern discovery: an inter-disciplinary survey. Wiley Interdisc. Rew.: Data Mining and Know. Discovery 4(1), 2014. (with X. Zhou et al.)

Algorithmic Fairness and Equity

Facial Recognition Is Accurate, if You're a White Guy Feb. 9, 2018 The New York Times

Color Matters in Computer Vision

Facial recognition algorithms made by Microsoft, IBM and Face++ were more likely to misidentify the gender of black women than white men.

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of

Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos.

Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Trust and Ethics: FATE debate

- **Government View:** Security, Balance prosperity and civil society
- **Business View:** Innovation critical for prosperity but carries risks
- **Civil Society View:** Risks should be disclosed
  - **F**airness (or equity) : Reduce bias across gender, race, age, …
  - **A**ccountability : Determine and assign responsibility for a machine judgement
  - **T**ransparency (or explainability): Be open and clear about (prediction) process
  - **E**thics:
    - Privacy-preserving, Use case specific dilemmas
    - Trustworhty: **S**afe (Do no harm), **S**ecure (Guard against malicious behavior)

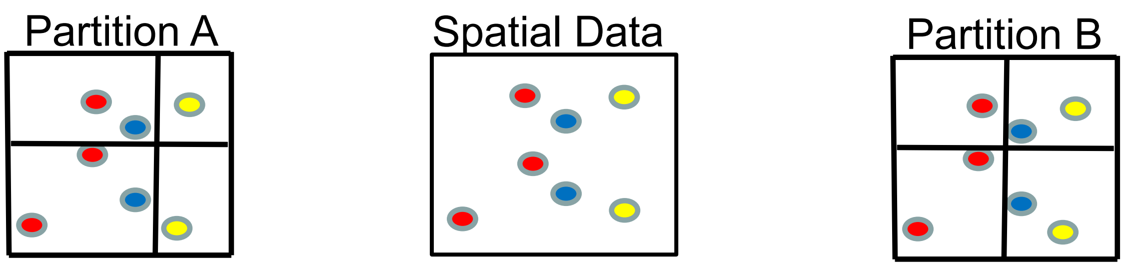
- **Q?** Which category does MAUP/Gerrymandering risk belong to? Choices: F, A, T, E

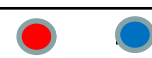
More: (i) Don't let industry write the rules for AI, Y. Benkler, Nature, 569, 161, 5/1/2019.
(ii) Data for Good: FATES, Elaborated, J. Wing, Jan. 23, 2018. (iii) FAT ML and FATES Workshop

https://www.fatml.org/

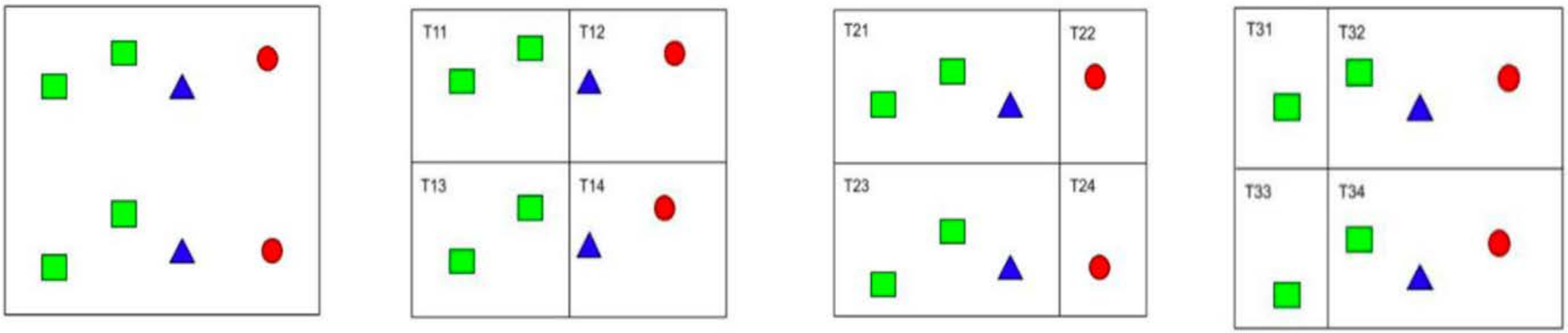
Gerrymandering Risk in Traditional Data Science

- Traditional methods not robust in face of
  - Spatial continuity
    - Gerrymandering risk: Spatial partitioning affects Results (Modifiable Areal Unit Problem)
  - Auto-correlation, Heterogeneity , Edge-effect, …
  - Noise challenge data mining methods



Partition A: Pearson's Correlation	Pairs	Partition B: Pearson's Correlation
1		- 0.90
- 0.90		1

Limitations of Traditional Data Mining: Association Rules



(a) Map of 3 item-types (b) Spatial Partition P1 (c) Spatial Partition P2 (d) Spatial Partition P3

Partitioning	**P1**	**P2**	**P3**
Transactions	T11, T12, T13, T14	T21, T22, T23, T24	T31, T32, T33, T44
Associations with support >= 0.5	( )	( )	( )