

# Spatial Data Science: Challenges & Opportunities

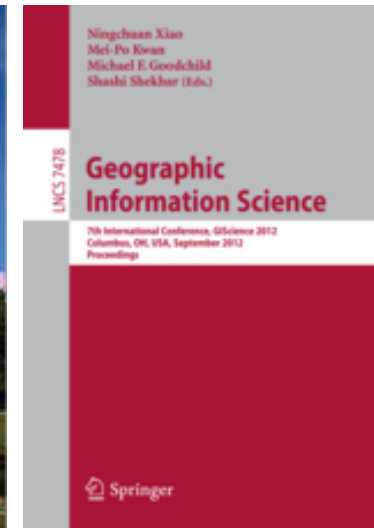
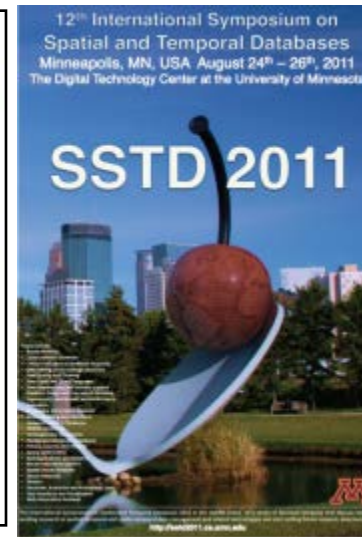
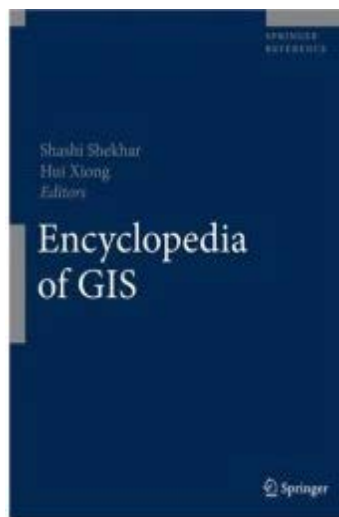
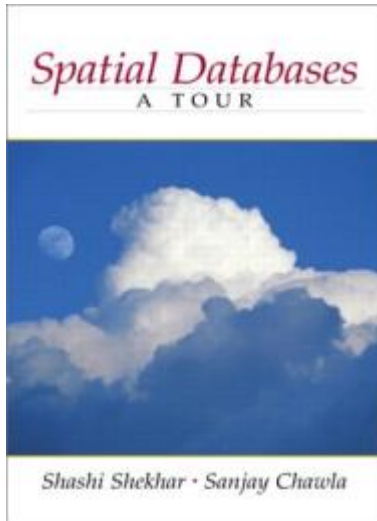
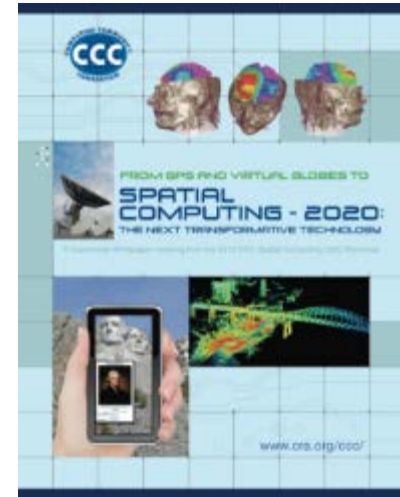
Shashi Shekhar

McKnight Distinguished University Professor

Department of Computer Science and Engineering

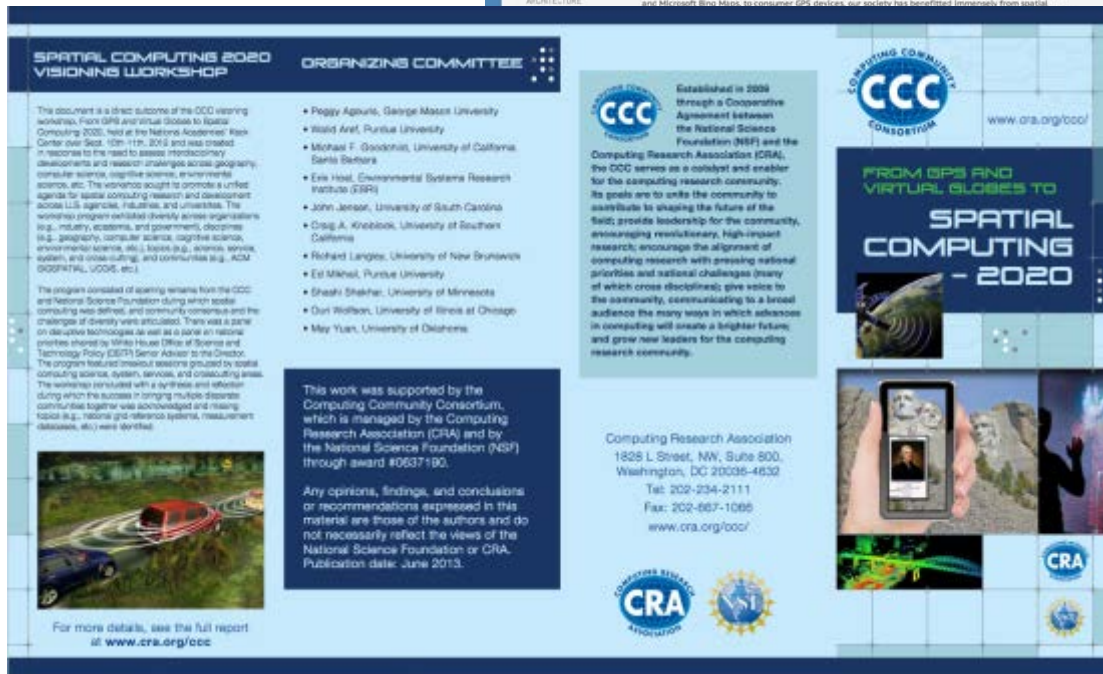
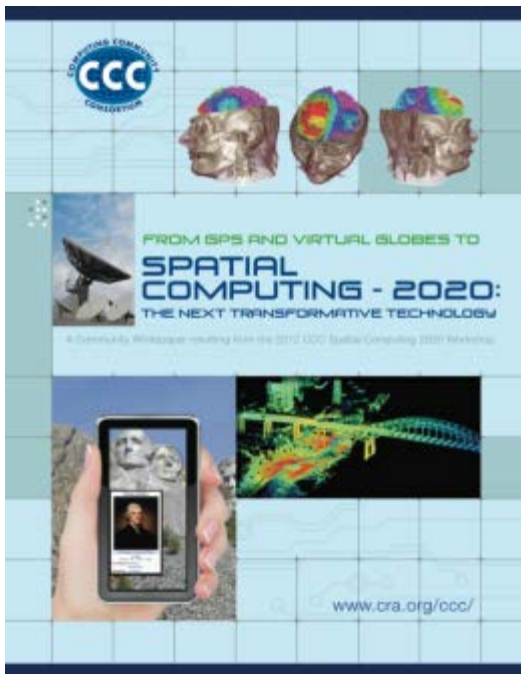
University of Minnesota

[www.cs.umn.edu/~shekhar](http://www.cs.umn.edu/~shekhar)



# Sources

- From GPS and Virtual Globes to Spatial Computing 2020, CCC Report, 2013. [www.cra.org/ccc/visionsing/visionsing-activities/spatial-computing](http://www.cra.org/ccc/visionsing/visionsing-activities/spatial-computing)
- With few slides on work from presenter's group  
Identifying patterns in spatial information: a survey of methods, Wiley Interdisc. Reviews: Data Mining and Know. Discovery , 1(3):193-214, May/June 2011. (DOI: 10.1002/widm.25).



# Outline



- Introduction
  - Spatial Computing Audience: Niche => Everyone
  - Spatial Computing 2020 - Workshop
- GPS
- Location Based Services
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes
- Geographic Information Systems
- Conclusions

# What is Spatial Computing?

- Transformed our lives through understanding spaces and places
  - Examples: localization, navigation, site selection, mapping,
  - Examples: spatial context, situation assessment (distribution, patterns), ...



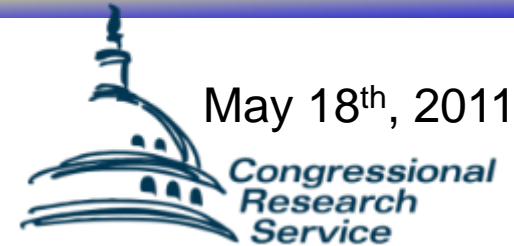


# The Changing World of Spatial Computing

<b>Late 20<sup>th</sup> Century</b>	<b>21<sup>st</sup> Century and Beyond</b>
Maps were produced by a few highly trained people in government agencies and surveying companies	Everyone is a mapmaker and many phenomena are observable.
Only sophisticated groups (e.g., Department of Defense, oil exploration groups) used GIS technologies	Everyone uses location-based services
Only specialized software (e.g., ArcGIS, Oracle SQL) could edit or analyze geographic information	Every platform is location aware
User expectations were modest (e.g., assist in producing and distributing paper maps and their electronic counterparts)	Rising expectations due to vast potential and risks

# It is widely used by Government!

## Geospatial Information and Geographic Information Systems (GIS): An Overview for Congress



**Table I. 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)

# It is only a start! Bigger Opportunities Ahead!

McKinsey Global Institute

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

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 study cites smartphone location services including Foursquare and Loopt, for locating friends, and ones for finding nearby stores and restaurants.

But 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

New Ways to Exploit Raw Data May Bring Surge of Innovation, a Study Says

# CCC Visioning Workshop: Making a Case for Spatial Computing 2020

[http://cra.org/ccc/spatial\\_computing.php](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 Participants

## Academia

## Industry

## Government

**Peggy Agouris**, George Mason University  
**Divyakant Agrawal**, University of California Santa Barbara  
**Cecilia Aragon**, University of Washington  
**Walid G. Aref**, Purdue University  
**Elisa Bertino**, Purdue University  
**Henrik Christensen**, Georgia Institute of Technology  
**Isabel Cruz**, University of Illinois at Chicago  
**Michael R. Evans**, University of Minnesota  
**Steven Feiner**, Columbia University  
**Jie Gao**, Stony Brook University  
**Michael Goodchild**, University of California Santa Barbara  
**Sara Graves**, University of Alabama Huntsville  
**Rajesh Gupta**, University of California San Diego  
**Chuck Hansen**, University of Utah  
**Stephen Hirtle**, University of Pittsburgh  
**Krzysztof Janowicz**, University of California Santa Barbara  
**John Jensen**, University of South Carolina  
**Daniel Keefe**, University of Minnesota  
**John Keyser**, Texas A&M University  
**Craig A. Knoblock**, Information Sciences Institute

**Hank Korth**, Lehigh University  
**Benjamin Kuipers**, University of Michigan  
**Vipin Kumar**, University of Minnesota  
**Richard Langley**, University of New Brunswick  
**Chang-Tien Lu**, Virginia Tech  
**Dinesh Manocha**, University of North Carolina  
**Edward M. Mikhail**, Purdue  
**Harvey Miller**, University of Utah  
**Joe Mundy**, Brown University  
**Dev Oliver**, University of Minnesota  
**Rahul Ramachandran**, UA Huntsville  
**Norman Sadeh**, CMU  
**Shashi Shekhar**, University of Minnesota  
**Daniel Z. Sui**, Ohio State  
**Roberto Tamassia**, Brown University  
**Paul Torrens**, University of Maryland  
**Shaowen Wang**, University of Illinois at Urbana-Champaign  
**Greg Welch**, University of North Carolina  
**Ouri E. Wolfson**, University of Illinois at Chicago  
**Mike Worboys**, University of Maine  
**May Yuan**, University of Oklahoma  
**Avidoh Zakhor**, University of California Berkeley

**Mark Abrams**, ESG  
**Mohamed Ali**, Microsoft  
**Lee Allison**, Arizona Geological Survey  
**Virginia Bacon Talati**, Computer Science and Telecommunications Board (CSTB)  
**Ramon Caceres**, AT&T Research  
**Vint Cerf**, Google  
**Jade DePalacios**, Naval Postgraduate School  
**Jon Eisenberg**, Computer Science and Telecommunications Board (CSTB)  
**Tom Erickson**, IBM  
**Erwin Gianchandani**, CCC  
**Eric Hoel**, ESPI  
**Xuan Liu**, IBM  
**Siva Ravada**, Oracle  
**Jagan Sankaranarayanan**, NEC Labs  
**Lea Shanley**, Wilson Center  
**Kevin Pomfret**, Centre for Spatial Law and Policy

**Nabil Adam**, DHS  
**Vijay Atluri**, NSF  
**David Balshaw**, NIH/NIEHS  
**Budhendra Bhaduri**, ORNL  
**Kelly Crews**, NSF  
**Beth Driver**, NGA  
**Walton Fehr**, USDOT  
**Myron Gutmann**, NSF  
**Susanne Hambrusch**, NSF  
**Michelle Heacock**, NIH/NIEHS  
**Clifford Jacobs**, NSF  
**Farnam Jahanian**, NSF  
**Todd Johanesen**, NGA  
**Thomas Johnson**, NGA  
**Henry Kelly**, OSTP  
**Alicia Lindauer**, USDOE  
**Keith Marzullo**, NSF  
**John L. Schnase**, NASA  
**Jim Shine**, Army Research  
**Raju Vatsavai**, ORNL  
**Eric Vessey**, NSA  
**Howard D. Wactlar**, NSF  
**Tandy Warnow**, NSF  
**Nicole Wayant**, Army Research  
**Mark Weiss**, NSF  
**Maria Zemankova**, NSF  
**Li Zhu**, NIH/NCI

>30 Universities

14 Organizations

12 Agencies

# 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



# Workshop Highlights

## Pull Panel: National Priorities, Societal Applications of Spatial Computing

**Chair:** Henry Kelly, OSTP

### Members

US-DoD: Eric Vessey

US-DoD: Todd Johanesen

NIH/NIEHS: Michelle Heacock

NASA: John L Schnase

DHS: Nabil Adam

NSF EarthCube: Clifford Jacobs

DOT: Walton Fehr

DOE: Alicia Lindauer

## Push Panel: Spatial Computing (SC) Platform Trends, Disruptive Technologies

**Chair:** Dinesh Manocha, UNC

### Members:

Graphics & Vision: John Keyser, TAMU

Interaction Devices: Steven Feiner, Columbia University

LIDAR : Avideh Zakhor, UCB

GPS Modernization: Mark Abrams, Advisor to USG

Cell Phones: Ramon Caceres, AT&T

Indoor Localization: Greg Welch, UNC

Internet Localization: Rajesh Gupta, UCSD

Cloud Computing: Divyakant Agarwal, UCSB

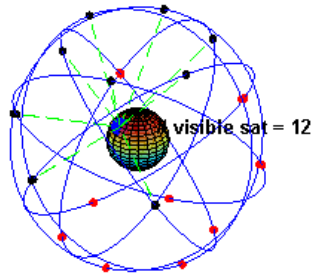
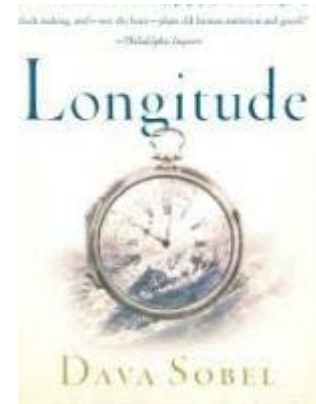
# Outline



- Introduction
- **GPS**
  - **Outdoors => Indoors**
- Location Based Services
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes
- Geographic Information Systems
- Conclusions

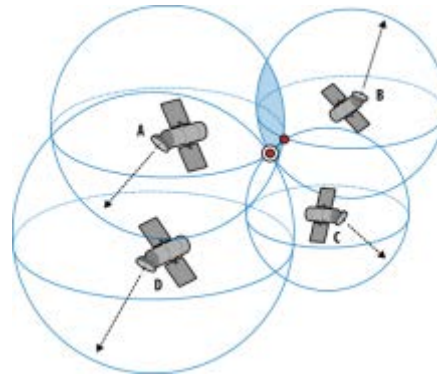
# Global Positioning Systems

- Positioning ships
  - Latitude (compass, star positions)
  - Longitude: dead-reckoning => marine chronometer
  - Longitude prize (1714), accuracy in nautical miles
- Global Positioning System (**GPS**), & competition
  - Infrastructure: satellites, ground stations, receivers, ...
  - Use: Positioning (sub-meter), Clock synchronization



[http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System)

Trilateration



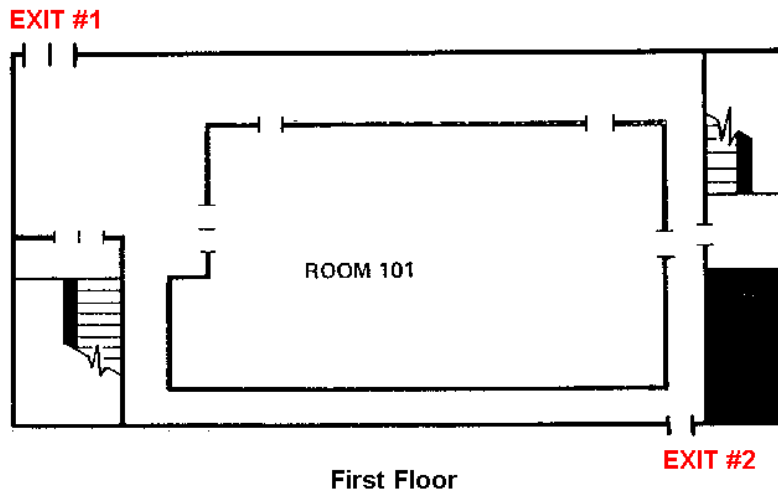
<http://answers.oreilly.com/topic/2815-how-devices-gather-location-information/>



# Trends: Localization Indoors and Underground

- GPS works outdoors, but,
  - We are indoors 90% of time!
  - Ex. malls, hospitals, airports, etc.
  - Indoor asset tracking, exposure hotposts, ...
- Leveraging existing indoor infrastructure
  - WiFi, Cell-towers, cameras, Other people?
- How to model indoors for navigation, tracking, hotspot analysis, ...?
  - What are nodes and edges ?

<http://www.mobilefringe.com/products/square-one-shopping-center-app-for-iphone-and-android/>



## WiFi Localization



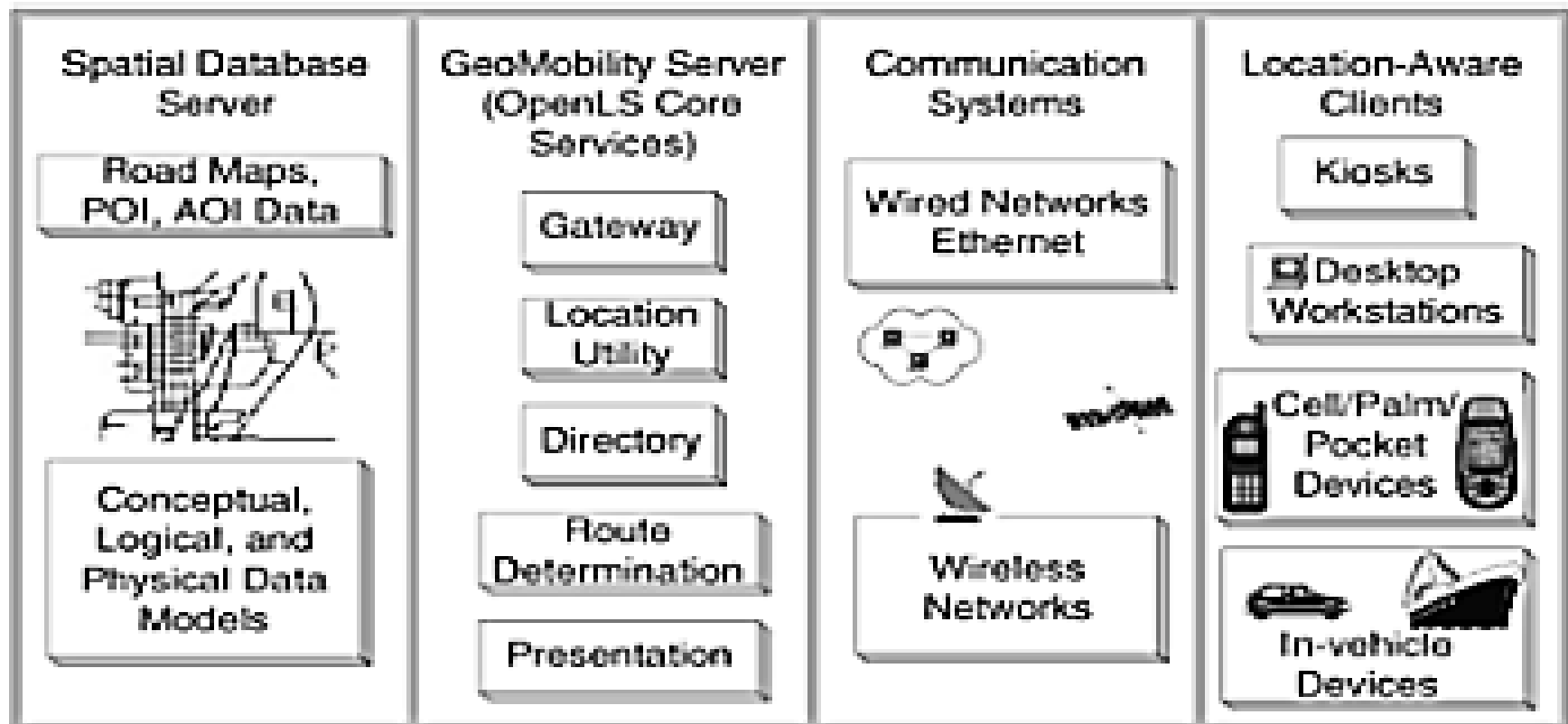
# Outline



- Introduction
- GPS
- Location Based Services
  - Queries => Persistent Monitoring
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes
- Geographic Information Systems
- Conclusions

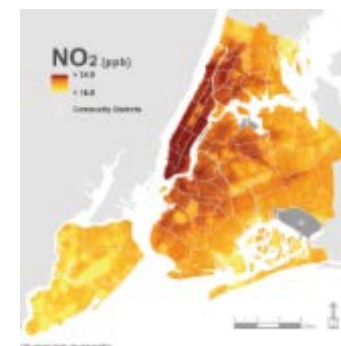
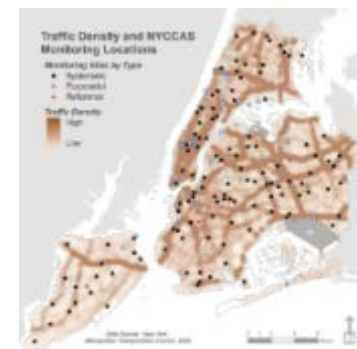
# Location Based Services

- Open Location Services: Queries
  - Location: Where am I? (street address, <latitude, longitude>
  - Directory: Where is the nearest clinic (or doctor)?
  - Routes: What is the shortest path to reach there?



# Trends: Persistent Environmental Hazard Monitoring

- Environmental influences on our health & safety
  - air we breathe, water we drink, food we eat
- Surveillance (e.g., SEER):
  - **Passive > Active > Persistent**
  - A fixed sensor covers a location, all the time
  - A moving sensor covers all location for some time!
  - **How may one economically cover all locations all the time ?**
- How do we create the infrastructure for the continuous and timely collection, fusion, curation and analysis of big spatio-temporal data?
  - Crowd-sourcing, e.g., American Redcross tweet maps
  - Smart-phone based sensors, e.g., NIEHS



# Outline

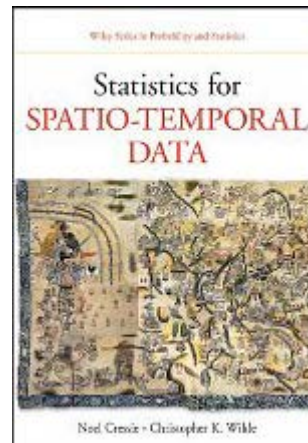
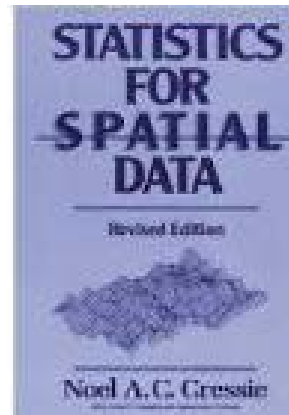
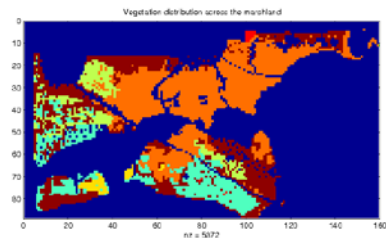
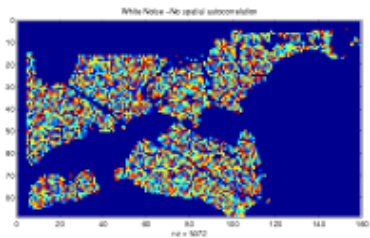
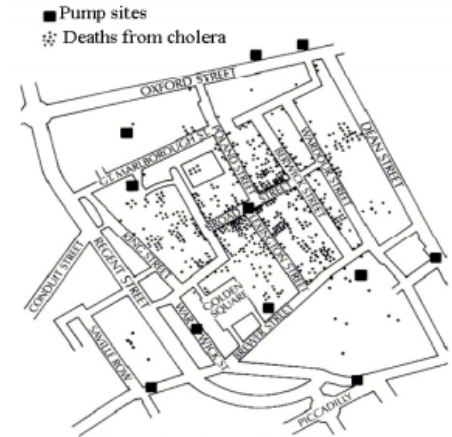


- Introduction
- GPS
- Location Based Services
- **Spatial Statistics**
  - **Concepts: Mathematical => Spatial**
- Spatial Database Management Systems
- Virtual Globes
- Geographic Information Systems
- Conclusions



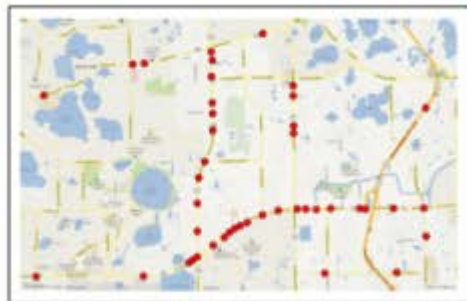
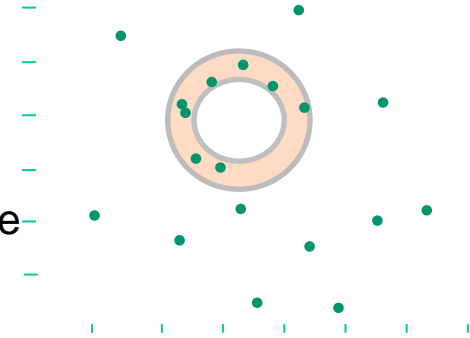
# Spatial Statistics: Mathematical Concepts

- Spatial Statistics
  - Quantify uncertainty, confidence, ...
  - Is it significant?
  - Is it different from a chance event or rest of dataset?
    - e.g., SaTScan finds circular hot-spots
- Models of Auto-correlation, Heterogeneity, Edge-effect, ...
  - Point Process, e.g., Ripley's K-functions, SatScan
  - Geo-statistics, e.g., Kriging, GWR
  - Lattice-based models



# Trends: Spatial-Concept Aware Patterns

- Spatial Concepts
  - Natural geographic features, e.g., rivers, streams, ...
  - Man-made geographic features, e.g., transportation network
  - Spatial theories, e.g., environmental criminology – doughnut hole
- Spatial-concept-aware patterns
  - Hotspots: Circle => Doughnut holes
  - Hot-spots => Hot Geographic-features



(a) Input



(b) Crimestat K-means with Euclidean Distance



(c) Crimestat K-means with Network Distance

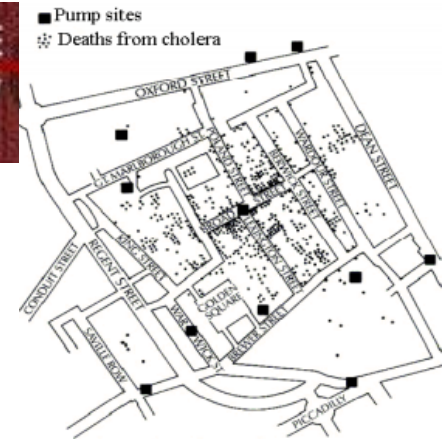


(d) KMR

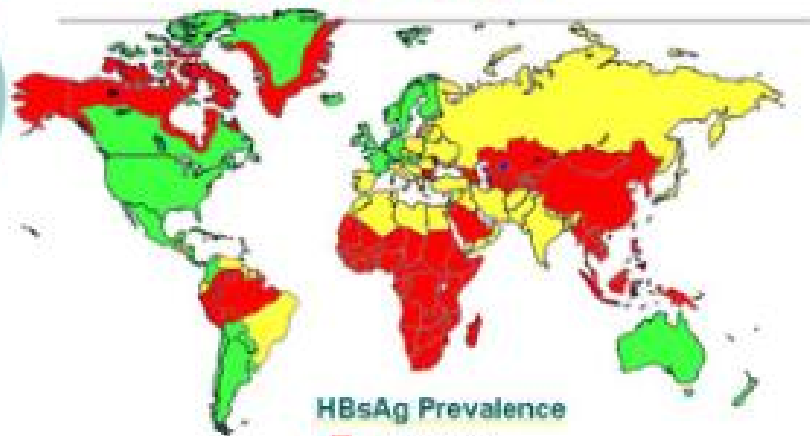
Source: Identifying patterns in spatial information: a survey of methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 193-214, 1(3), May/June 2011. (DOI: 10.1002/widm.25).

# Hypothesis generation via Co-locations/Co-occurrence

- Co-location (Cholera Deaths, Water Pump)
  - Hypothesis: Cholera is water-borne (1854)
  - Miasma theory => Germ Theory
- Co-location (Liver Cancer, HBV infection)
- Which exposures and cancers are co-located?
  - How may we find these co-location in big data?



## Geographic Distribution of Chronic HBV Infection

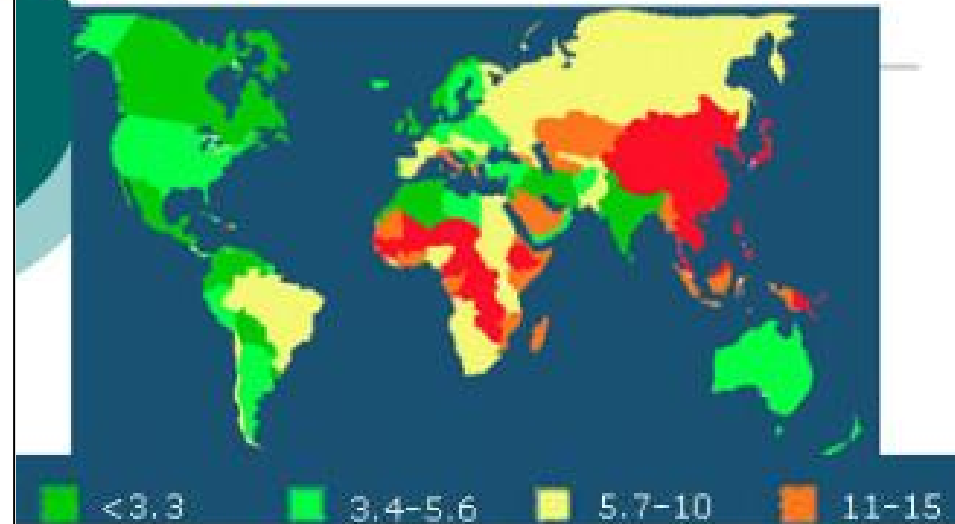


HBsAg Prevalence

- $\geq 8\%$  - High
- 2-7% - Intermediate
- $< 2\%$  - Low

© CDC 2003

## Incidence of Liver Cancer in Males- 2002



$< 3.3$

3.4-5.6

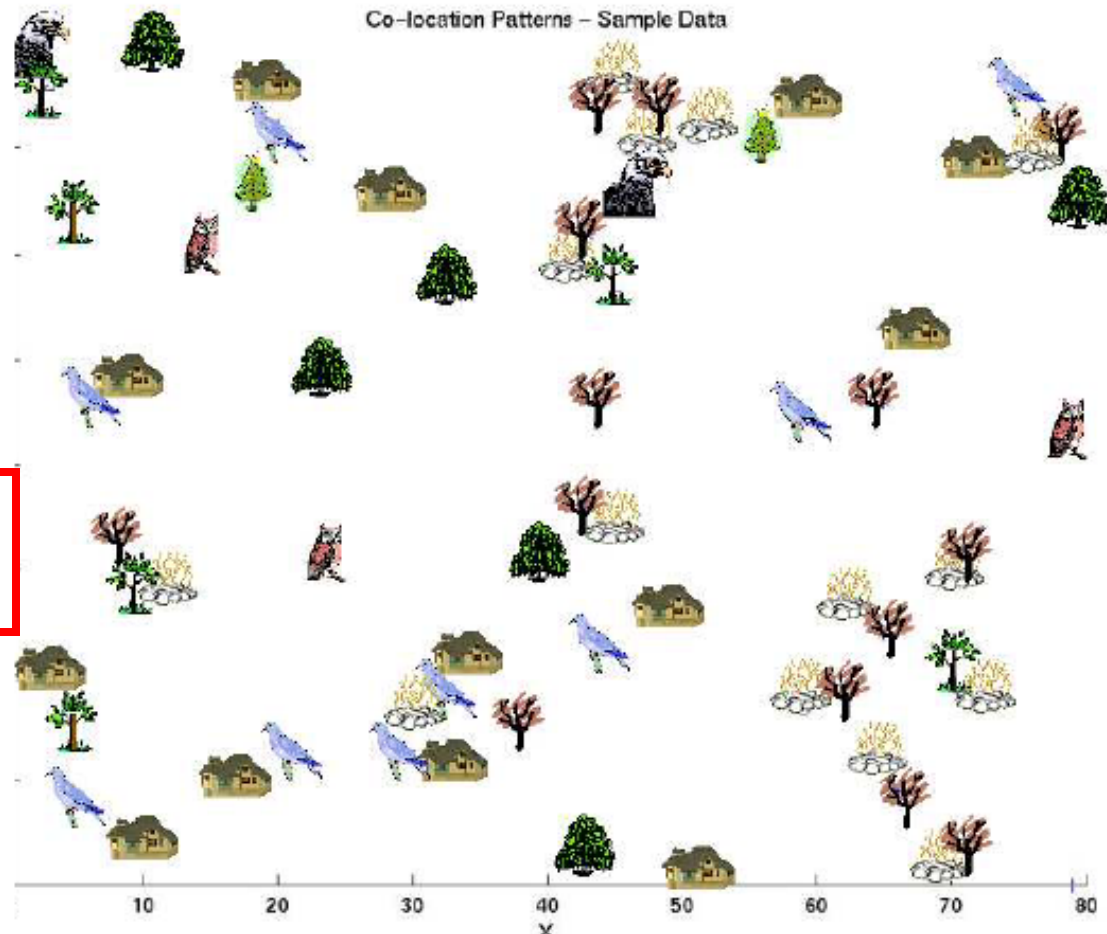
5.7-10

11-15

Rates per 100,000

# Co-locations/Co-occurrence

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



Source: Identifying patterns in spatial information: a survey of methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery , 193-214, 1(3), May/June 2011. (DOI: 10.1002/widm.25).

# Fast Algorithms to Mine Colocations from Big Data

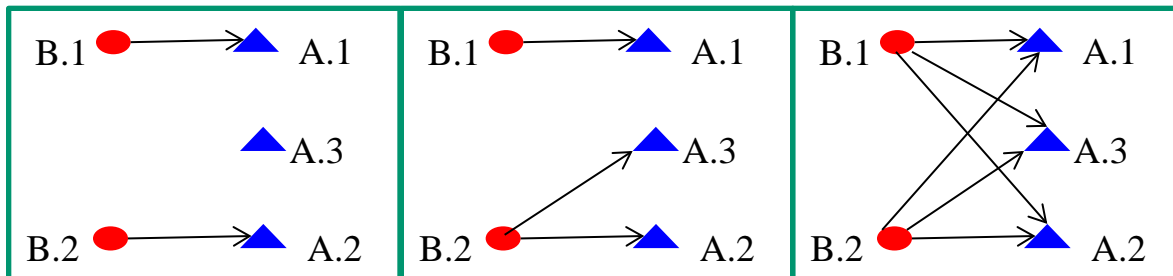
**Participation ratio**  $pr(f_i, c)$  of feature  $f_i$  in colocation  $c = \{f_1, f_2, \dots, f_k\}$ :  
 fraction of instances of  $f_i$  with feature  $\{f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k\}$  nearby  
 (i.e. within a given distance)

**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

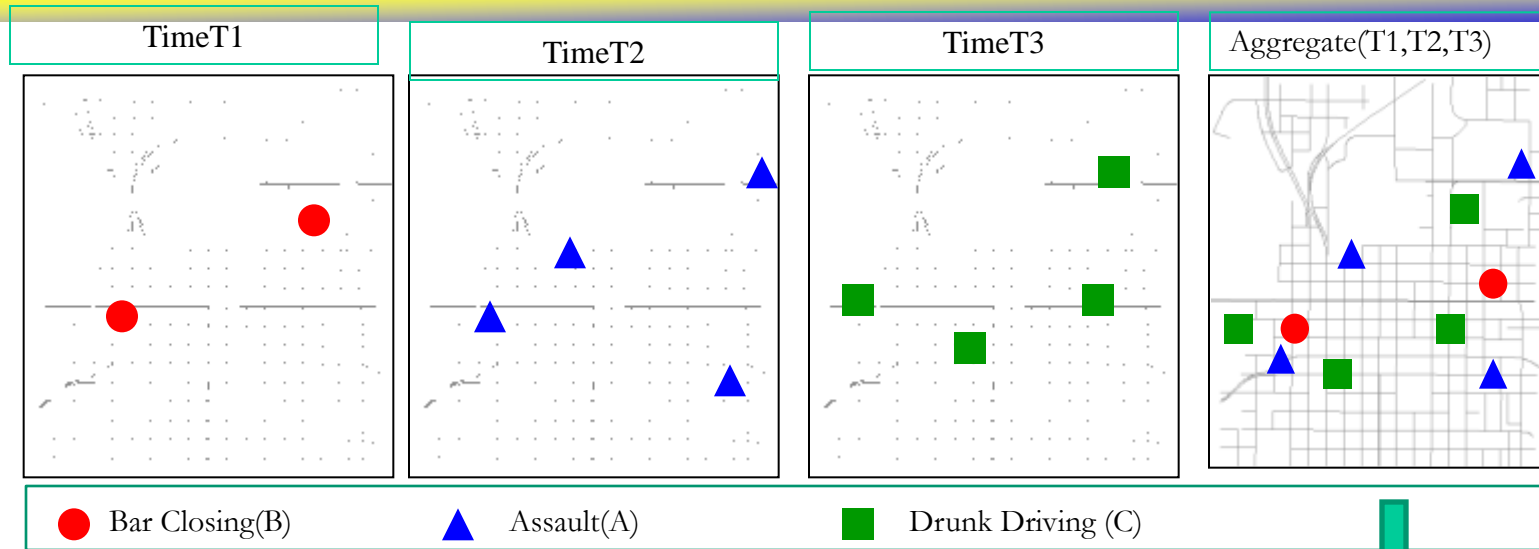
## ■ Comparison with Ripley's K-function (Spatial Statistics)



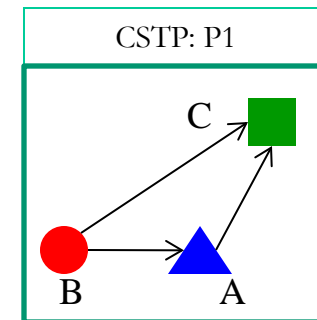
<b>K-function (B → A)</b>	$2/6 = 0.33$	$3/6 = 0.5$	$6/6 = 1$
<b>PI (B → A)</b>	$2/3 = 0.66$	1	1



# 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, ...



Source: Identifying patterns in spatial information: a survey of methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery , 193-214, 1(3), May/June 2011. (DOI: 10.1002/widm.25).

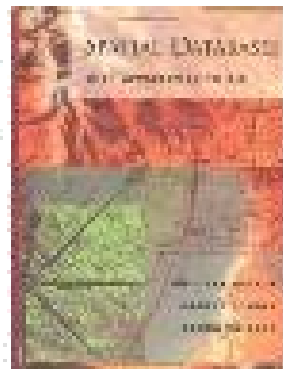
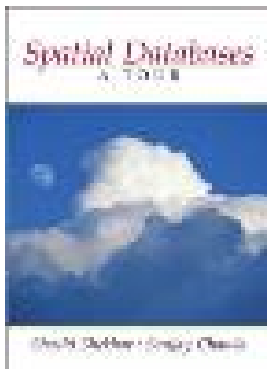
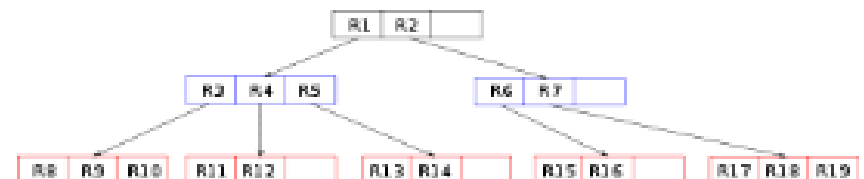
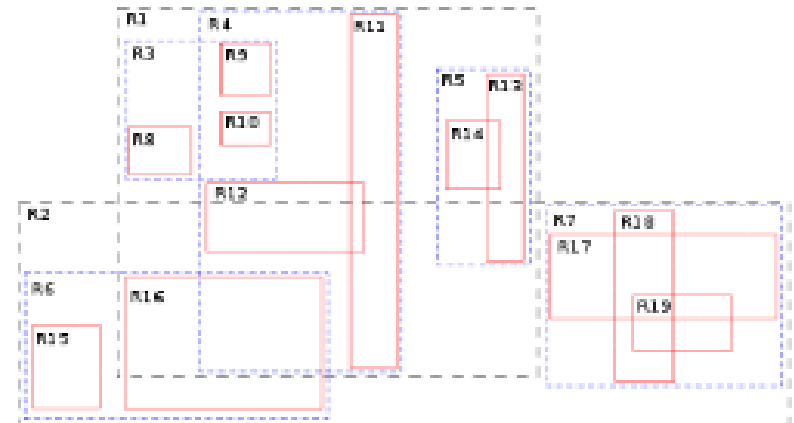
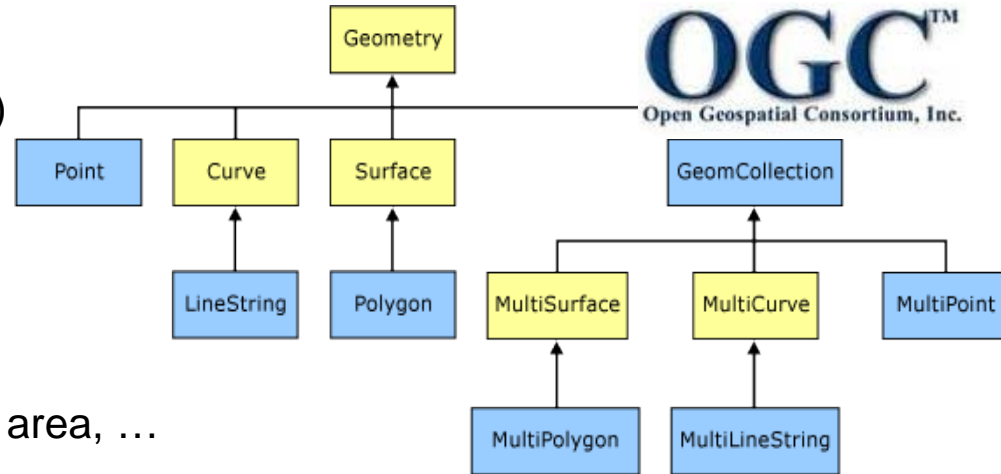
# Outline



- Introduction
- GPS
- Location Based Services
- Spatial Statistics
- **Spatial Database Management Systems**
  - Geometry => Spatial Network Databases
- Virtual Globes
- Geographic Information Systems
- Conclusions

# Spatial Databases for Geometry

- Dice, Slide, Drill-down, Explore, ...
  - Closest pair( school, pollution-source)
  - Set based querying
- Reduce Semantic Gap
  - Clumsy code for inside, distance, ...
  - 6 data-types
  - Operations: inside, overlap, distance, area, ...
- Scale up Performance
  - Data-structures: B-tree => R-tree
  - Algorithms: Sorting => Geometric



# Trend: Spatial Network Databases

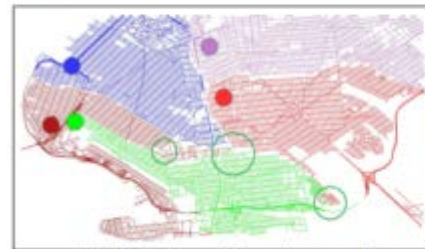
- Motivation: Urban phenomena
  - Access to medical facilities
  - Disparity in access
- Trend: model facility capacities
  - Graph models to Flow-Networks
- Trend: Network Patterns
  - (Origin, Destination) pairs
  - => Frequent Routes
- Trend: Spatio-temporal Networks



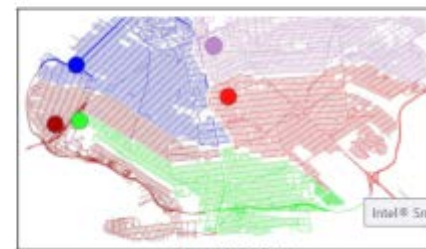
(a) Brooklyn Road Network and Gas Stations



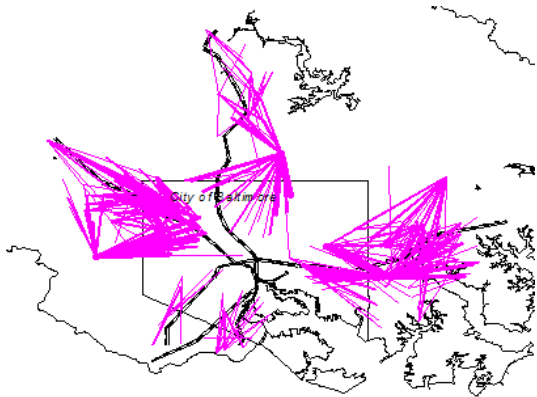
(b) NVD



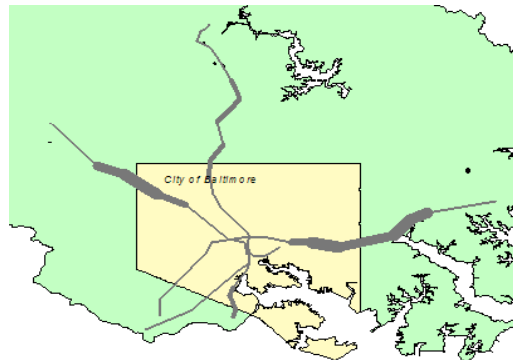
(c) Min-Cost Flow Approach



(d) CCNVD



(a) Input: Pink lines connect crime location & criminal's residence



(b) Output: *Journey-to-Crime*  
(thickness = route popularity)  
Source: Crimestat

# Challenge: Geo-privacy, geo-confidentiality, ...

- Emerging personal geo-data
  - Trajectories of smart phones, gps-devices, life-trajectories and migrations, ...
- Privacy: Who gets my data? Who do they give it to? What promises do I get?
- Socio-technical problem
  - Need government support
  - Challenges in fitting location privacy into existing privacy constructs (i.e HIPPA, Gramm-Leach-Bliley, Children's Online Privacy Protection Act)
- Groups interested in Geo-Privacy
  - Civil Society, Economic Entities, Public Safety ,Policy Makers

**Table 4.2: Geo-privacy Policy Conversation Starters**

1. Emergencies are different (E-911)
2. Differential geo-privacy can improve safety (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

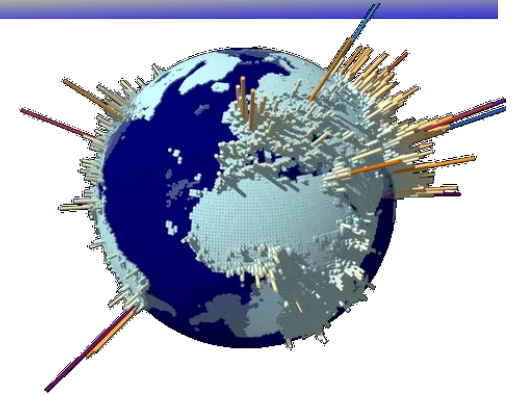


- Introduction
- GPS
- Location Based Services
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes & VGI
  - Quilt => Time-travel & Depth
- Geographic Information Systems
- Conclusions

# Virtual Globes & Volunteered Geo-Information

- Virtual Globes

- Visualize Spatial Distributions, Patterns
- Visual drill-down, e.g., fly-through
  - Change viewing angle and position
  - Even with detailed Streetview!



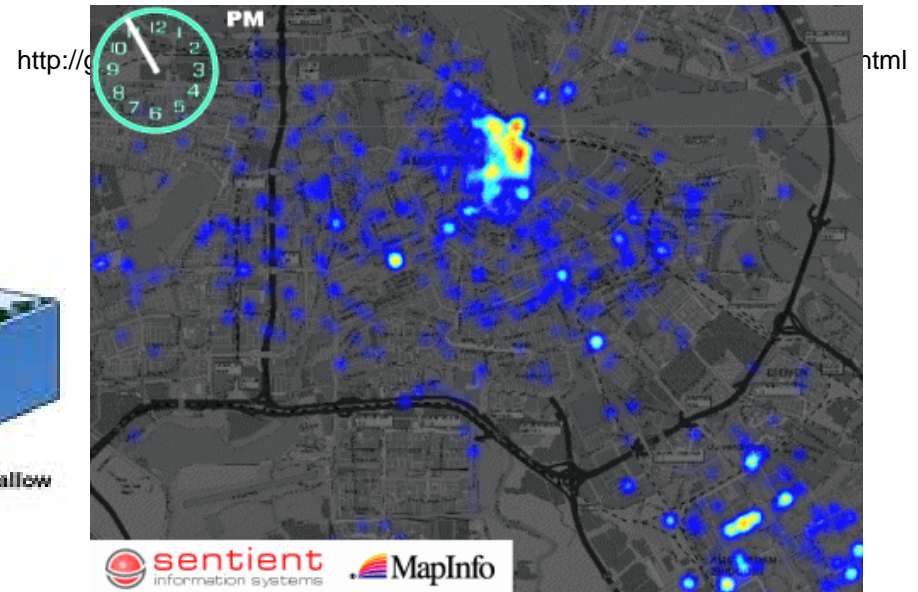
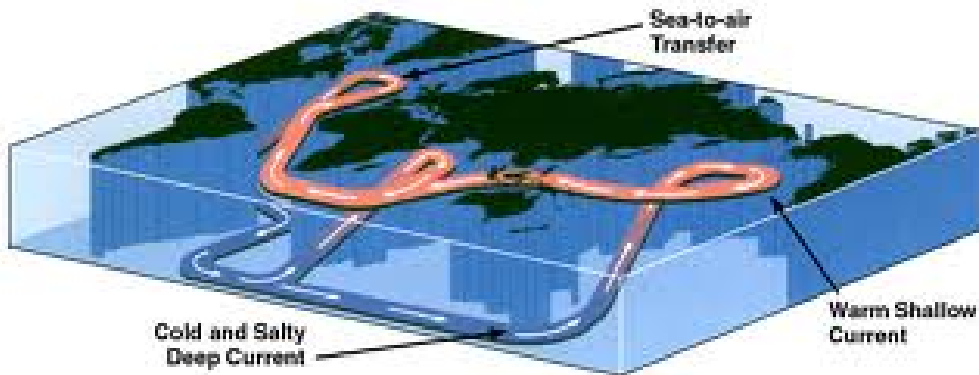
- Volunteered Geo-Information

- Allow citizens to make maps & report
- Coming to public health!
- People's reporting registry (E. Brokovich)
- [www.brockovich.com/the-peoples-reporting-registry-map/](http://www.brockovich.com/the-peoples-reporting-registry-map/)



# Opportunities: Time-Travel and Depth in Virtual Globes

- Virtual globes are snapshots
- How to add time? depth?
  - Ex. Google Timelapse: 260,000 CPU core-hours for global 29-frame video
- How may one convey provenance, accuracy, age, and data semantics?
- What techniques are needed to integrate and reason about diverse available sources?



# Outline



- Introduction
- GPS
- Location Based Services
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes
- **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.
- **Reference Systems**
  - Which countries in North Korea missile range?
  - 3D Earth surface displayed on 2D plane
  - Spherical coordinates vs. its planar projections



## North Korea's missiles

At least 1,000 of various types, according to South Korea's defense ministry



### Key arsenal

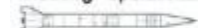
**Taepodong-2** First successful launch December 12, 2012 (Unha-3 rocket based on same system)



**Taepodong-1** Tested 1998 (failed)



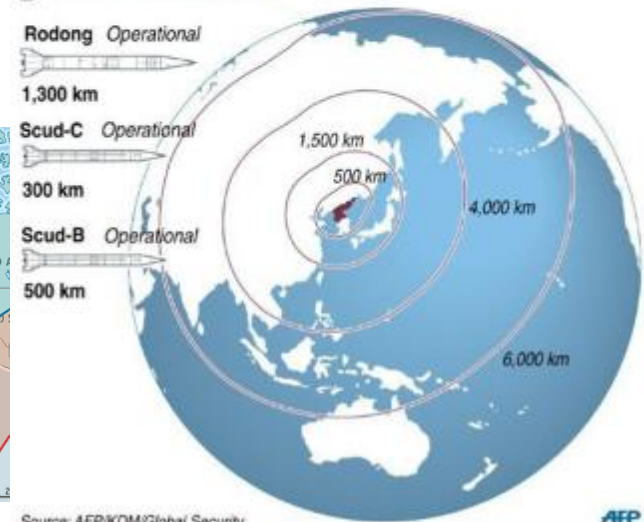
**Rodong** Operational



**Scud-C** Operational



**Scud-B** Operational



Source: AFP/KDM/Global Security

AFP

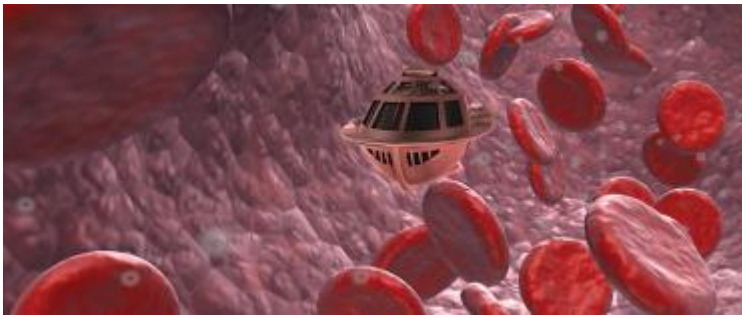


The Economist

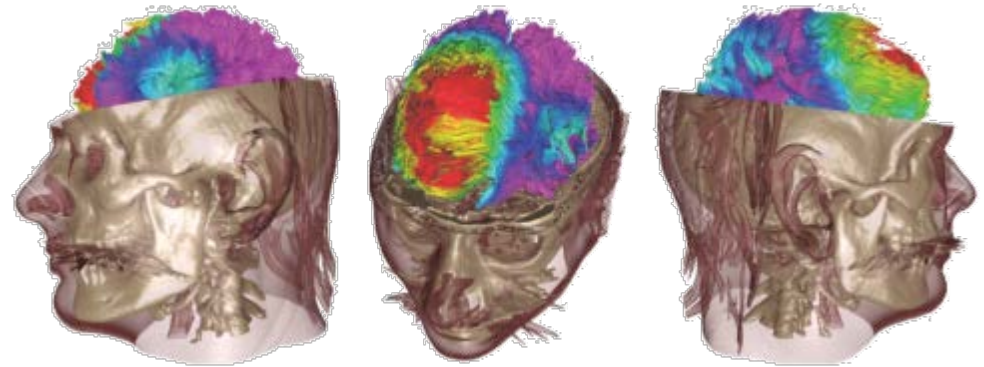
# Opportunities: Beyond Geographic Space

- 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>





# Outline



- Introduction
- GPS
- Location Based Services
- Spatial Statistics
- Spatial Database Management Systems
- Virtual Globes
- Geographic Information Systems
- **Conclusions**

# Recommendations

- Spatial Computing has transformed our society
  - It is only a beginning!
  - It promises an astonishing array of opportunities in coming decade
- However, these will not materialize without support
- Universities
  - Institutionalize spatial computing
    - GIS Centers, a la Computing Centers of the 1960's
  - Incorporate spatial thinking in STEM curriculum
    - During K-12, For all college STEM students?
- Government
  - Support spatial computing projects
  - Large and adventurous projects
    - 12+ faculty across multiple universities

