

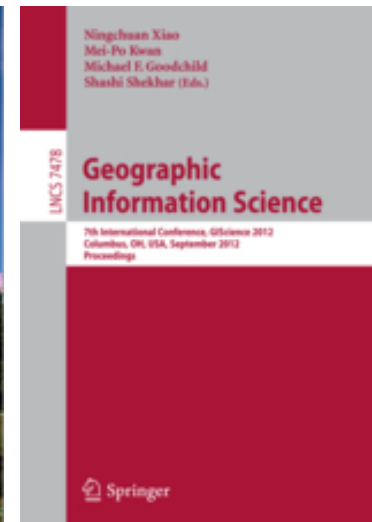
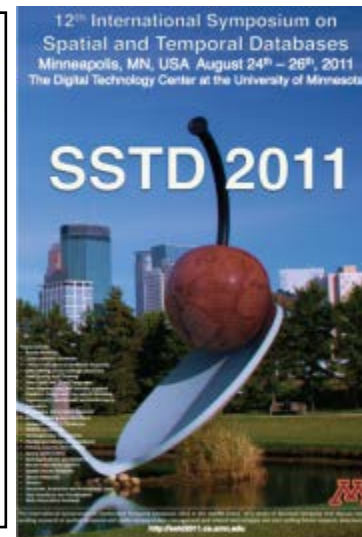
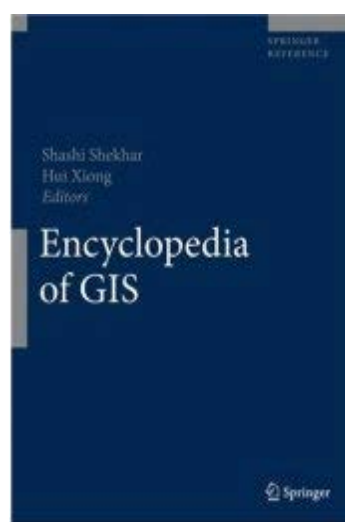
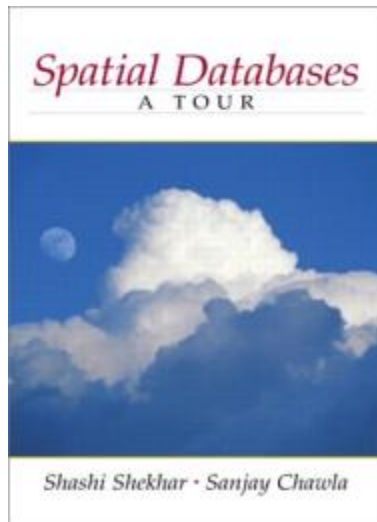
# Geo-spatial Approaches to Cancer Control:

## From GPS, Google Maps to Spatial Data Science

Sept. 2017

**Shashi Shekhar**

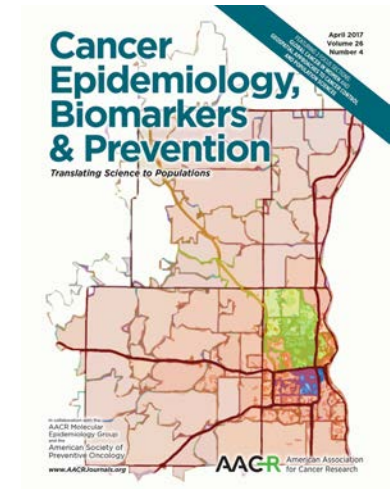
McKnight Distinguished University Professor  
Computer Sc. and Eng., University of Minnesota  
[www.cs.umn.edu/~shekhar](http://www.cs.umn.edu/~shekhar)



Thanks: Sponsors: NSF, NIH, USDOD, NASA, USDOT, USDOE, Ford URP, IBM, Microsoft, ...  
UMN: **Logan Spector** (MCC), CTS, IonE, DTC, Spatial Computing Group

# Outline

- Introduction
  - Spatial Computing Audience: Niche => Everyone
  - NCI Workshop : Geospatial Approaches to Cancer ...
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Statistics
  - Spatial Database Management Systems
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions



# Spatial Context Matters



**Francis S. Collins** ✓

@NIHDirector



Follow

T. Glass: If DNA is our biological blueprint **ZNA (zipcode at birth)** is the blueprint for behavioral&psycho-social makeup.

[#PMINetwork](#)

10:10 AM - May 29, 2015

However, precision medicine can't just look gene-deep. **Where we live –** the air we breathe, the water we drink, the environments around us – has a huge impact on our health and even on our DNA.

**THE CONVERSATION**  
Academic rigor, journalistic flair

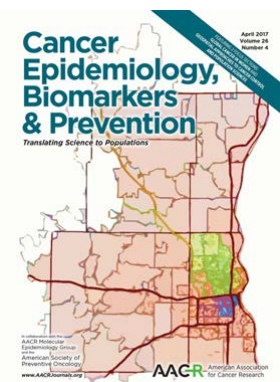


SCHOOL OF PUBLIC HEALTH

**Zip code better predictor of health than genetic code**  
[Melody Goodman](#)

# Geospatial Approaches to Cancer Control

- Special Issue, Cancer Epi., Biomarkers & Prevention, April 2017.
  - Follow the North Star: Why Space, **Place**, and Power Matter for Geospatial Approaches to Cancer Control and **Health Equity**
  - Visualizing the **Diffusion** of Digital Mammography in New York State
  - **GPS**-Based **Exposure** to Greenness and Walkability and Accelerometry-Based Physical Activity
  - Racial/Ethnic Differences in the **Impact of Neighborhood** Social and Built Environment on Breast Cancer Risk: The Neighborhoods and Breast Cancer Study
  - Spatial Analysis of **Regional Factors** and Lung Cancer Mortality in China, 1973–2013
  - Exploring the Linkage between **Activity-Friendly Zoning**, Inactivity, and Cancer Incidence in the United States
  - Housing Discrimination, Residential Racial **Segregation**, and Colorectal Cancer Survival in Southeastern Wisconsin
  - ...



# Conference on Geospatial Approaches to Cancer Control and Population Sciences

September 12-14, 2016

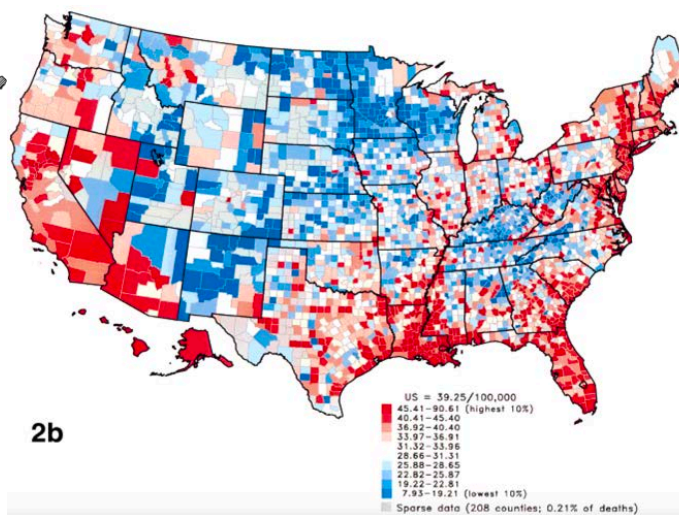
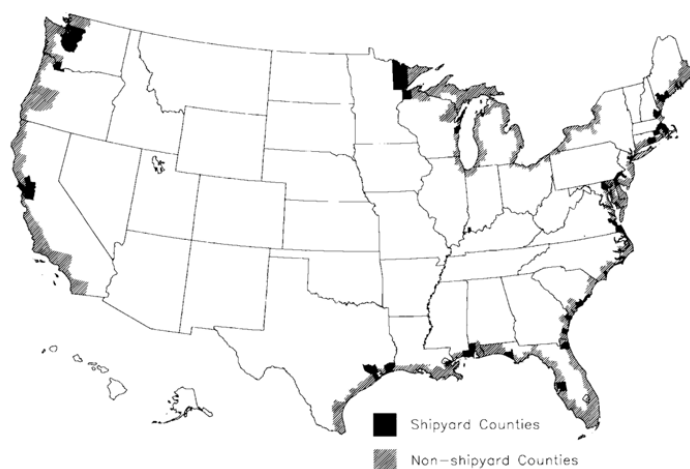
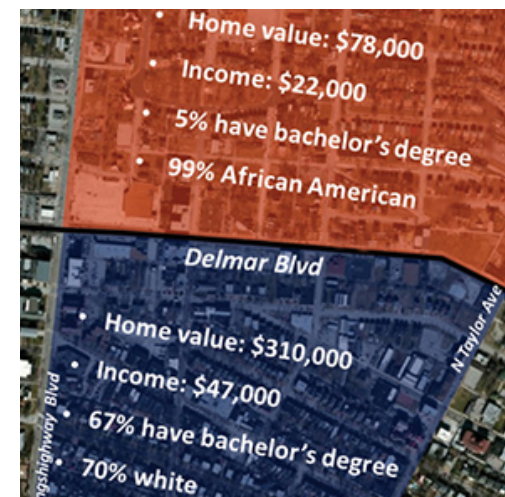
Natcher Conference Center, NIH Campus, Bethesda MD

- Keynotes
  - Segregation and the People's Health: Implications for Cancer Registries, Research, and Prevention (Nancy Krieger, Harvard)
  - Cancer Control and Populations from a Geospatial Perspective: A Focus on **NCI-Cancer Center Catchments** ( Tracy Onega, Dartmouth)
- Sessions
  1. Defining Personal Environments for Risk
  2. Geography of Screening and Vaccine
  3. Cancer Health **Disparities** and the Neighborhood and Social Environments
  4. Identifying **Priority (Geographic) Areas** for Cancer Control Activities
  5. **Accessibility** to Health Services
  6. Physical Environment and Cancer Risk
  7. Geography of **Health Care Delivery**
  8. Social Environment and Cancer Risk
  9. Geo-Visualization of Cancer Burden
  10. **Geo-Surveillance** of Cancer
  11. Geo-Statistical Methods and Models for Cancer Control
  12. Spatial Analyses Using **Surveillance** and Health Systems Data
  13. Geoinformatics, e.g., geocoding, ...
  14. Confidentiality and Defining Place Without Compromising Confidentiality



# Why Spatial Context Matters for Cancer?

- Built and Social Environments
  - Neighborhood affects access to healthy life-style
  - Ex. Delmar divide, St. Louis, MO
- Access to Health Services
  - Screening, Treatment
  - Ex. Appalachia Ohio
- Environmental Exposure
  - WW2 ship building, increased incidence of lung cancer



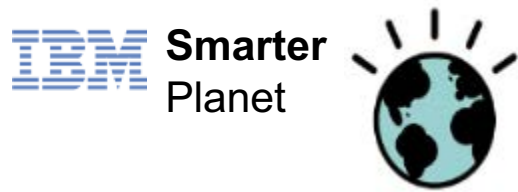
# Outline

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



# What is Spatial Computing?

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



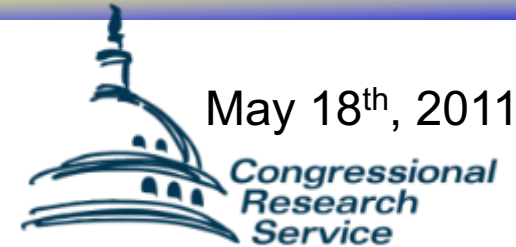


# The Changing World of Spatial Computing

	Last Century	Last Decade
Map User	Well-trained few	
Mappers	Well-trained few	
Software, Hardware	Few layers, e.g., Applications: Arc/GIS, Databases: SQL3/OGIS	
User Expectations & Risks	Modest	

# It is widely used by Government!

Q? Which agencies sowed seeds for Google Maps?



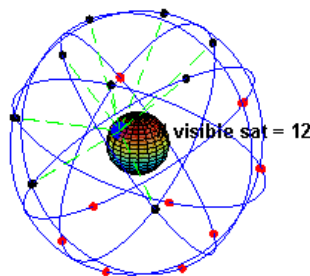
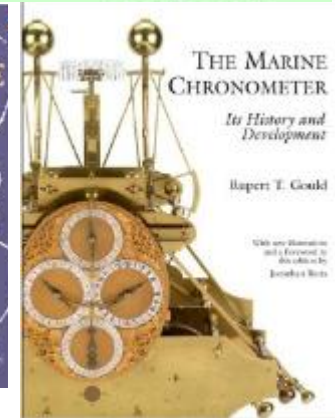
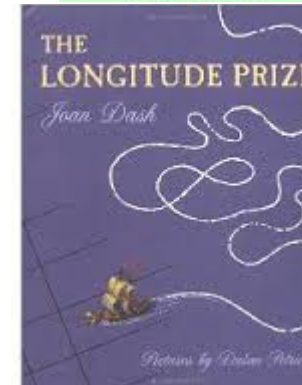
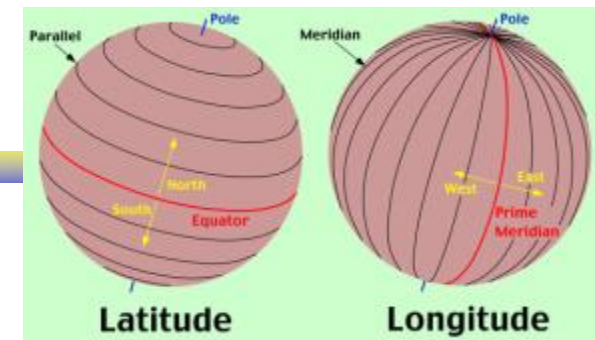
**Table 1. Members of the Federal Geographic Data Committee (FGDC)**

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)

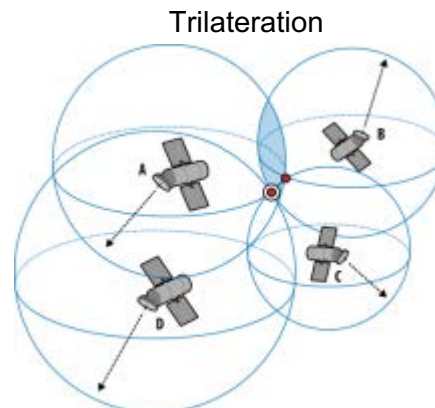
**Source:** Peter Folger, Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges. Congressional Research Service. June 8<sup>th</sup>, 2009. 10

# Global Positioning Systems (GPS)

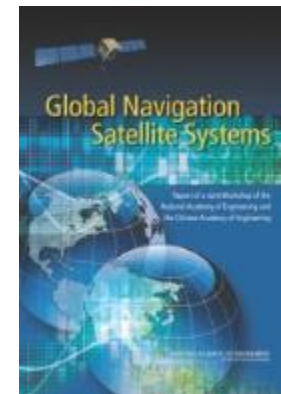
- Positioning ships
  - Latitude (compass, star positions)
  - Longitude Prize (1714) => marine chronometer
  - accuracy in nautical miles
- Global Navigation Satellite Systems
  - Use: Positioning, Clock synchronization
  - Infrastructure: satellites, ground stations, receivers, ...

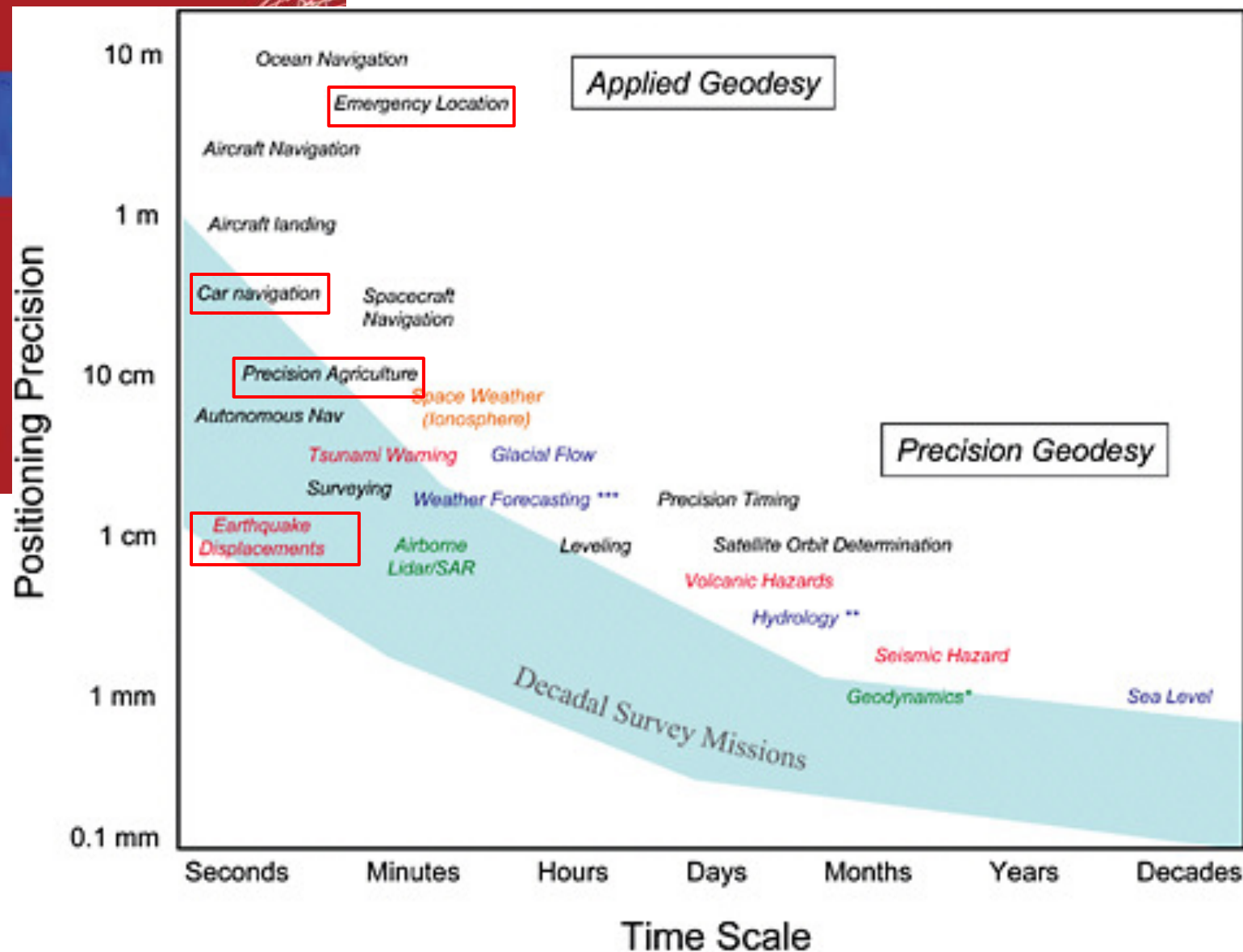


[http://en.wikipedia.org/wiki/Global\\_Positioning\\_System](http://en.wikipedia.org/wiki/Global_Positioning_System)



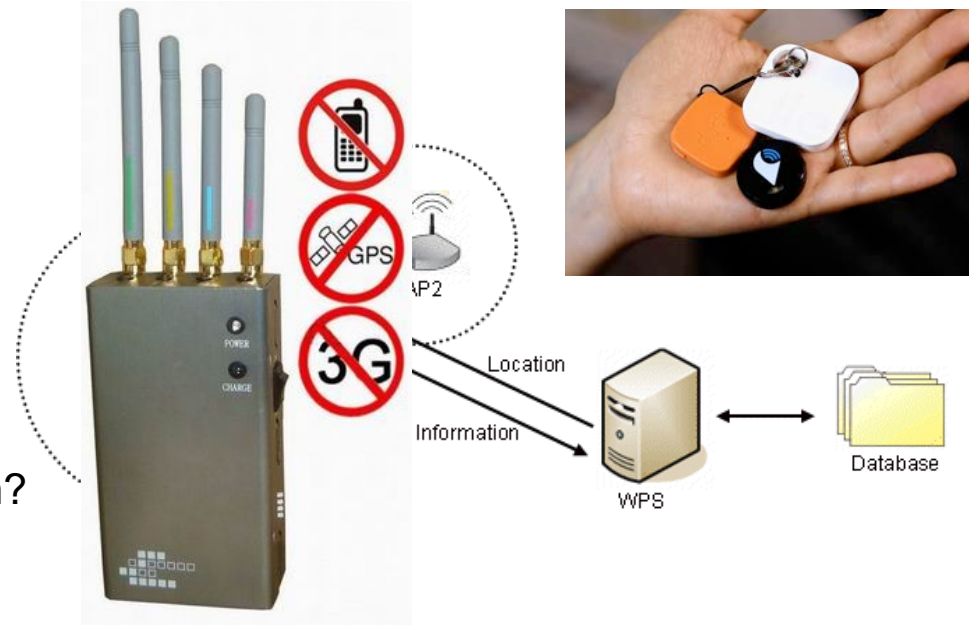
<http://answers.oreilly.com/topic/2815-how-devices-gather-location-information/>





# 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?



## TOP 10 LOCATION BASED SERVICES AT AIRPORTS

- |                          |                                 |
|--------------------------|---------------------------------|
| #1 FIND YOUR GATE        | #6 RECOMMENDED ACTIVITIES       |
| #2 YOUR CURRENT LOCATION | #7 PEOPLE FLOW OPTIMISATION     |
| #3 FIND [ANY SERVICE]    | #8 LOCATION BASED NOTIFICATIONS |





# 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 16<sup>th</sup>, 2012). <sup>14</sup>

# Outline



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

# Location Based Services

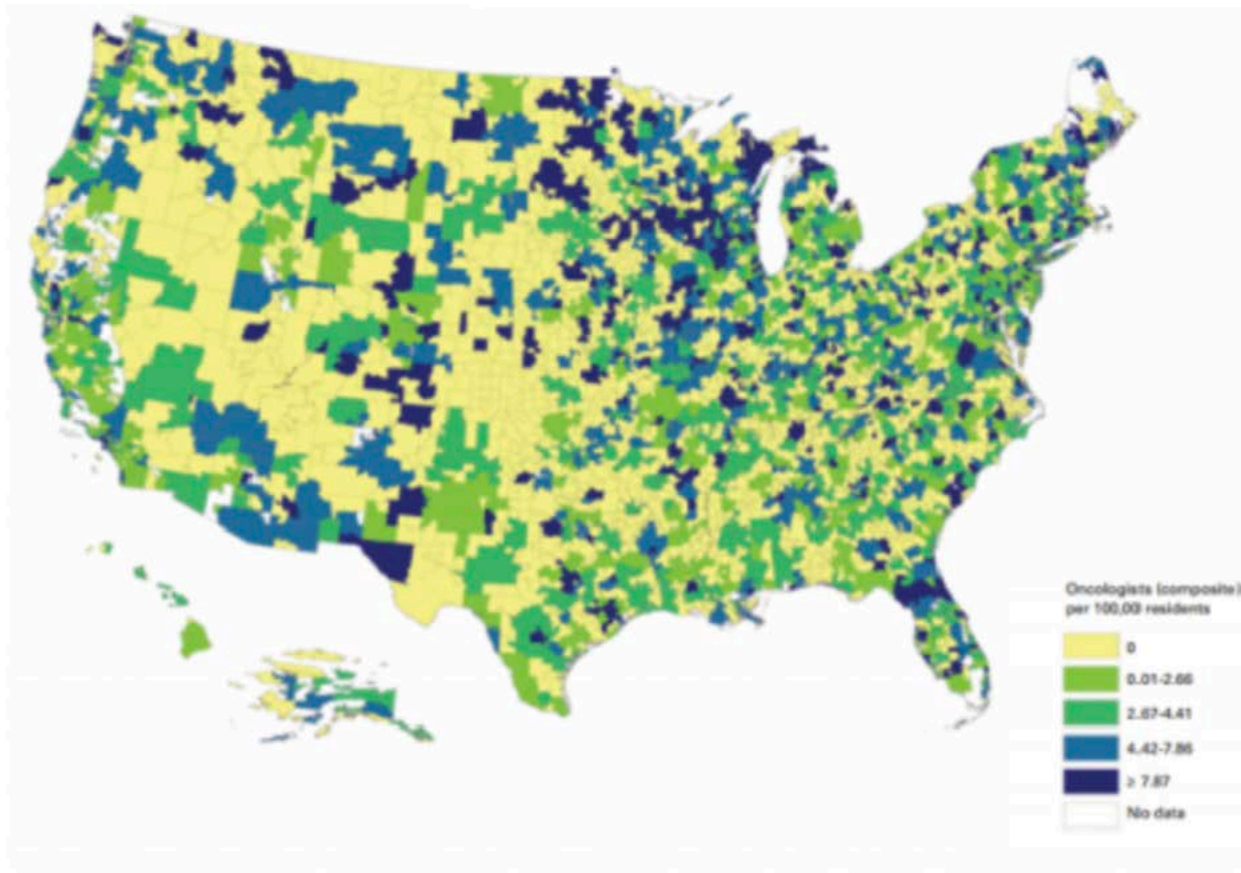
- Location: Where am I? (street address, <latitude, longitude>)
- Directory:
  - What is around me?
  - Where is the nearest clinic (or ambulance)?
- Routes: What is the shortest path to reach there?



# Access to Care

*Association Between Geographic Access to Cancer Care, Insurance, and Receipt of Chemotherapy: Geographic Distribution of Oncologists and Travel Distance*

*Chun Chieh Lin, et. al. JCO. October 2015.*



*Oncologists per 100,000 residents by hospital service area.*

# 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/](http://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



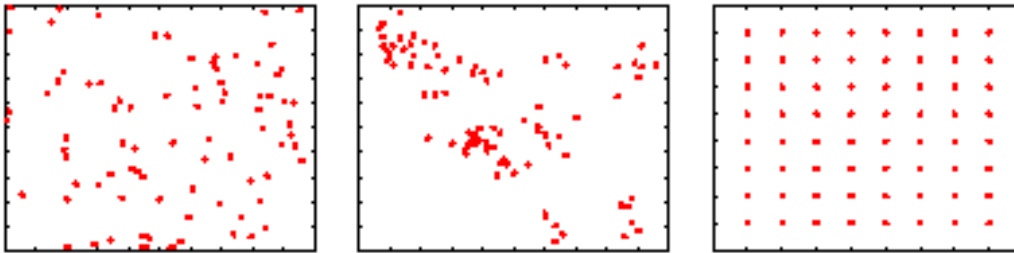
# Outline



- Introduction
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Statistics
    - From Mathematical (e.g., hotspot)
    - To Spatial (e.g., hot features)
  - Spatial Database Management Systems
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions

# Limitation of Traditional Clustering

- Simulation, Statistics, Data Mining, Machine Learning
- Challenge: **One size does not fit all**
  - Prediction error vs. model bias, Cost of false positives, ...
- Ex. Clustering: Find groups of tuples

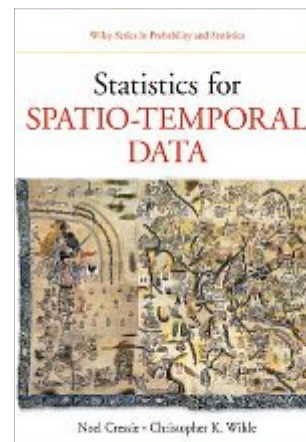
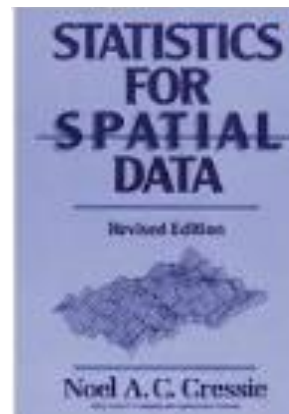
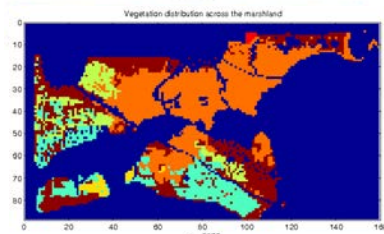
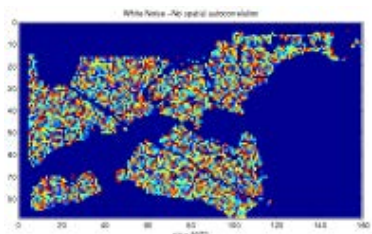
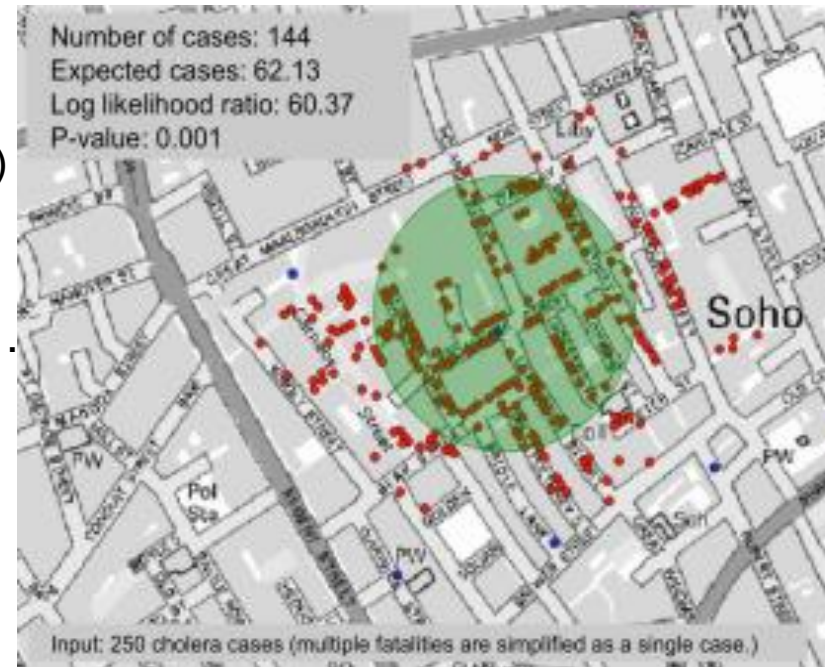


Traditional Clustering  
(K-means always finds clusters)

Spatial Clustering begs to differ!

# Spatial Statistics: Mathematical Concepts

- Spatial Statistics
  - Quantify uncertainty, confidence, ...
  - ?Different from chance events (rest of dataset)
    - e.g., SaTScan finds circular hot-spots
    - Tests for statistical significance
- Auto-correlation, Heterogeneity, Edge-effect, ...
  - Point Process, e.g., Ripley's K, SatScan
  - Geo-statistics, e.g., Kriging, GWR
  - Lattice-based models



# 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

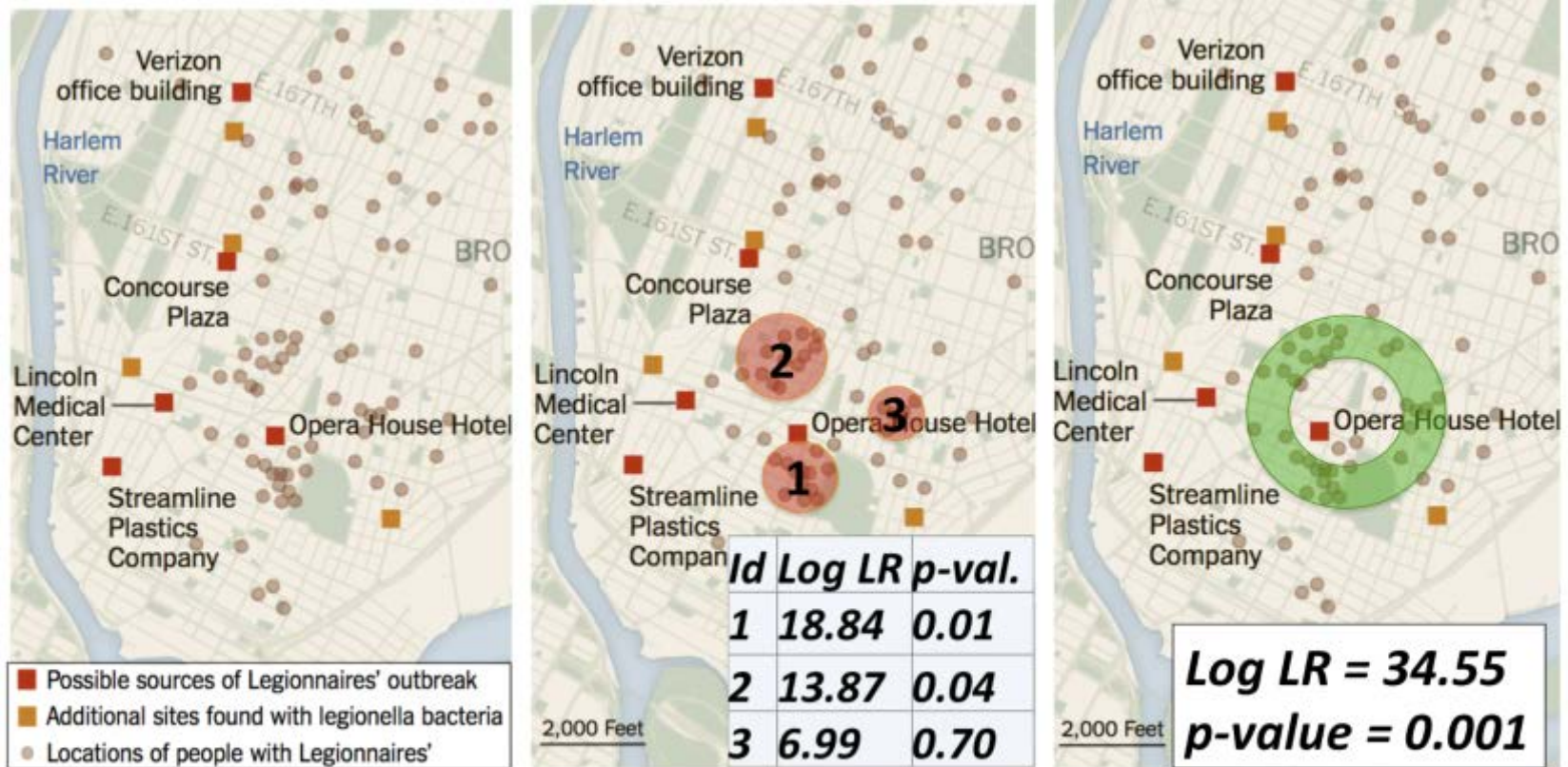
By The New York Times



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



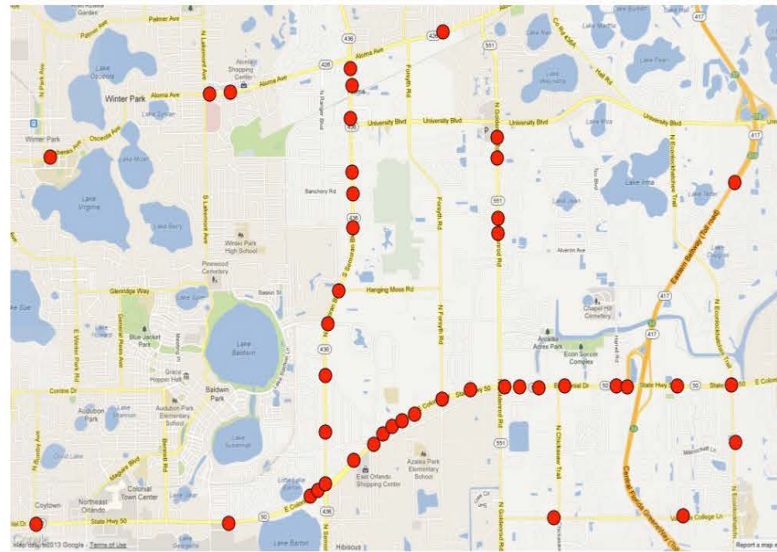
(a) Legionnaire's in New York (2015) (b) Output of SaTScan (c) Output of RHD

Source: Ring-Shaped Hotspot Detection, IEEE Trans. Know. & Data Eng., 28(12), 2016.

(A Summary in Proc. IEEE ICDM 2014) (w/ E. Eftelioglu et al.)



# Linear hotspots



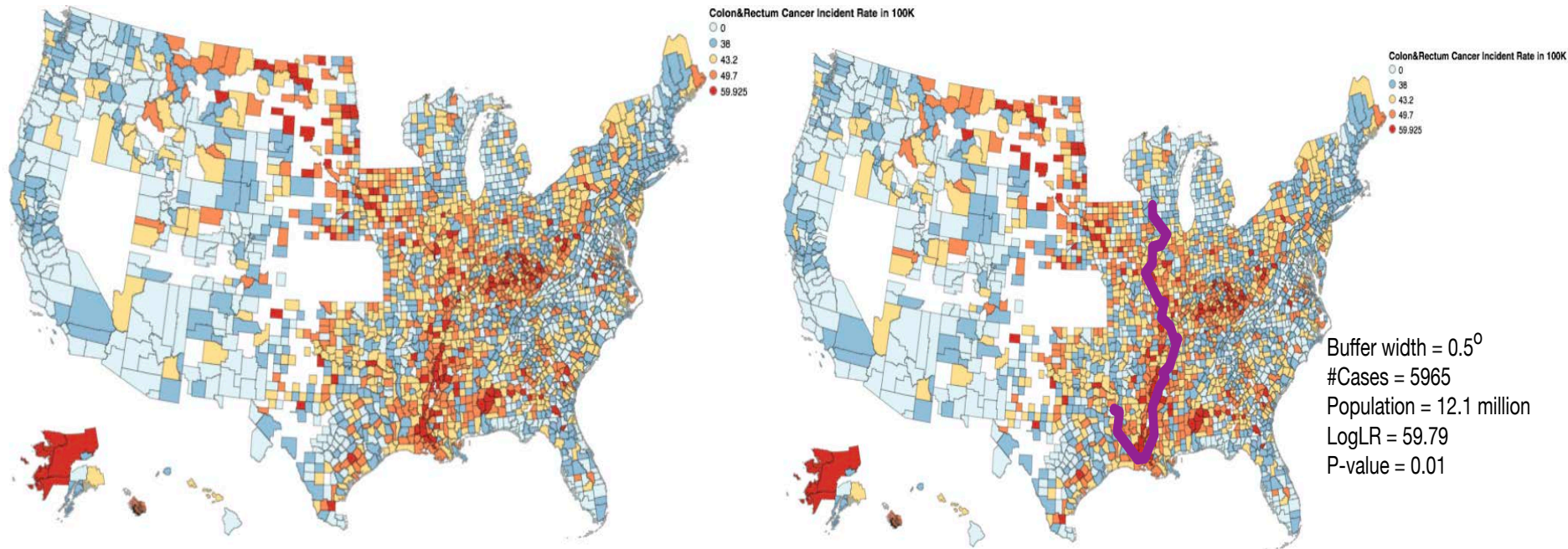
Data:  
43 pedestrian fatalities  
occurred in Orlando, FL

Circular hotspots by  
SatScan

Linear hotspots

Details: [Significant Route Discovery: A Summary of Results](#), Proc. Intl. Conf. on Geographic Information Science, Springer LNCS 8728, pp. 284-300, 2014.

# Linear Hotspot: Colo-rectum cancer incidence rate



**Included Counties in Continental U.S.: 2624**

**Population (2011): 297,511,048**

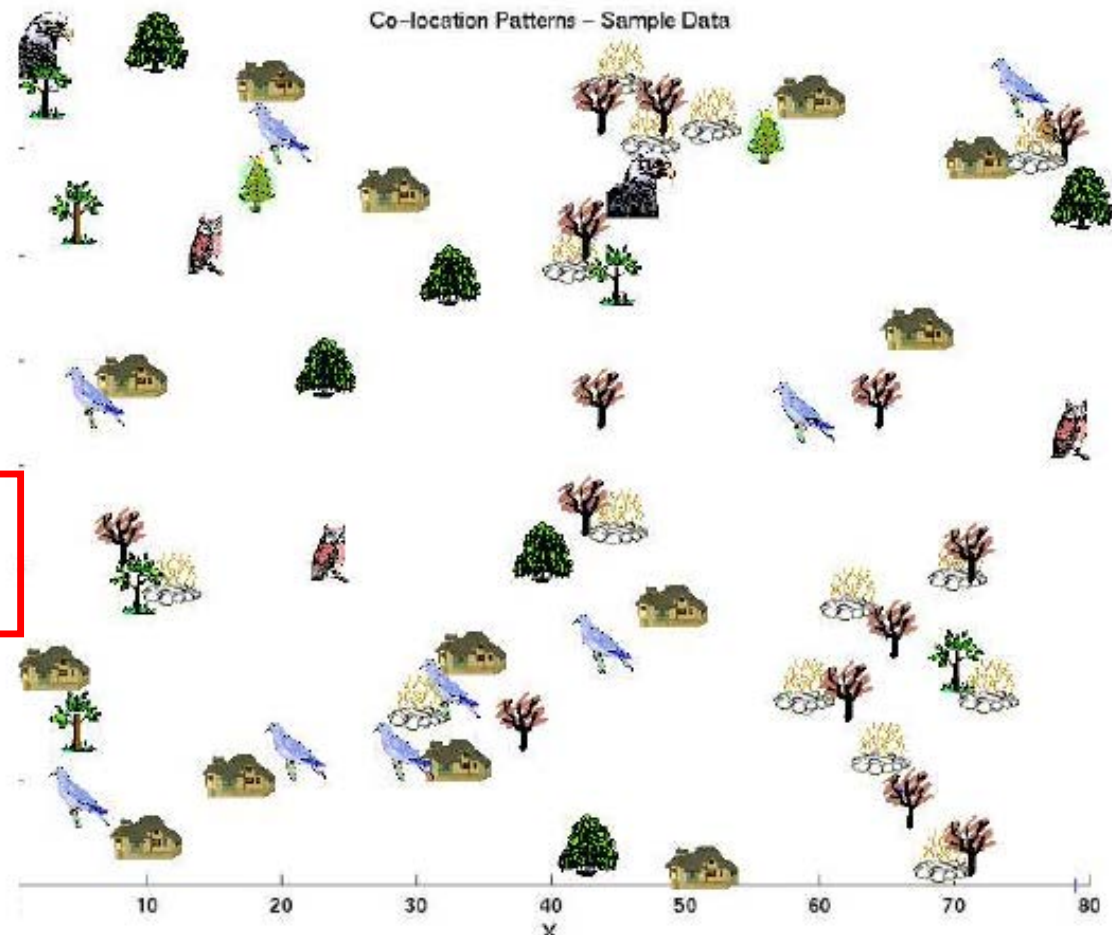
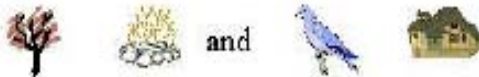
**Cancer Deaths (2009-2013): 120,572**

**Data Source: NCI website**

# Co-locations/Co-occurrence

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

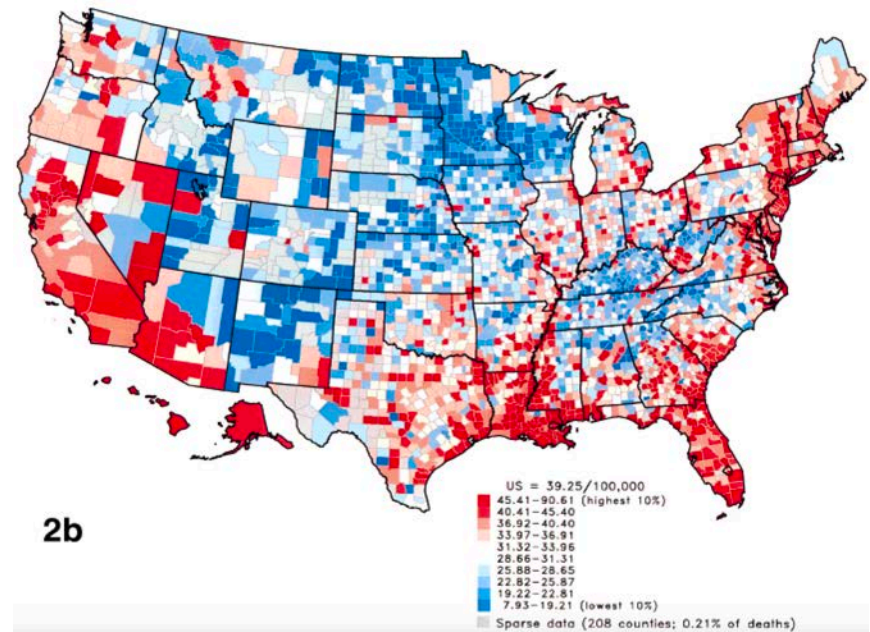
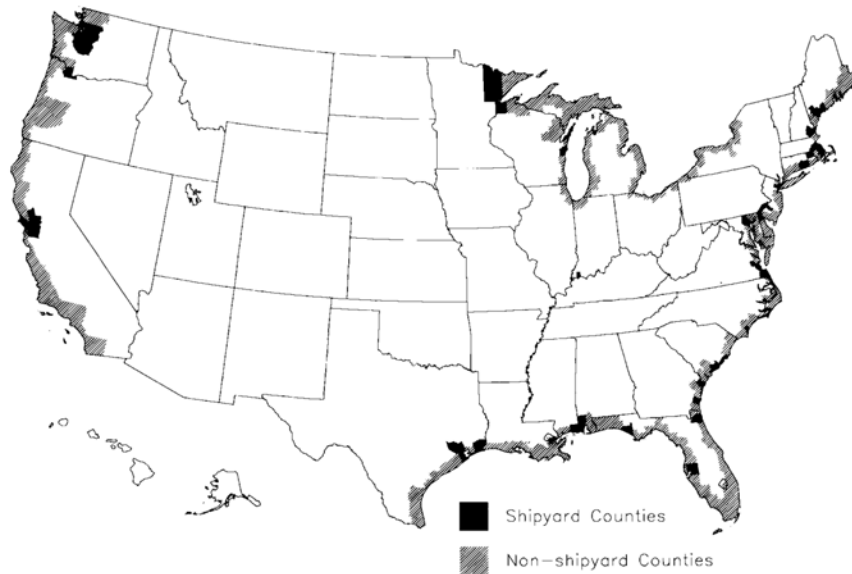
Answers:



**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 Example

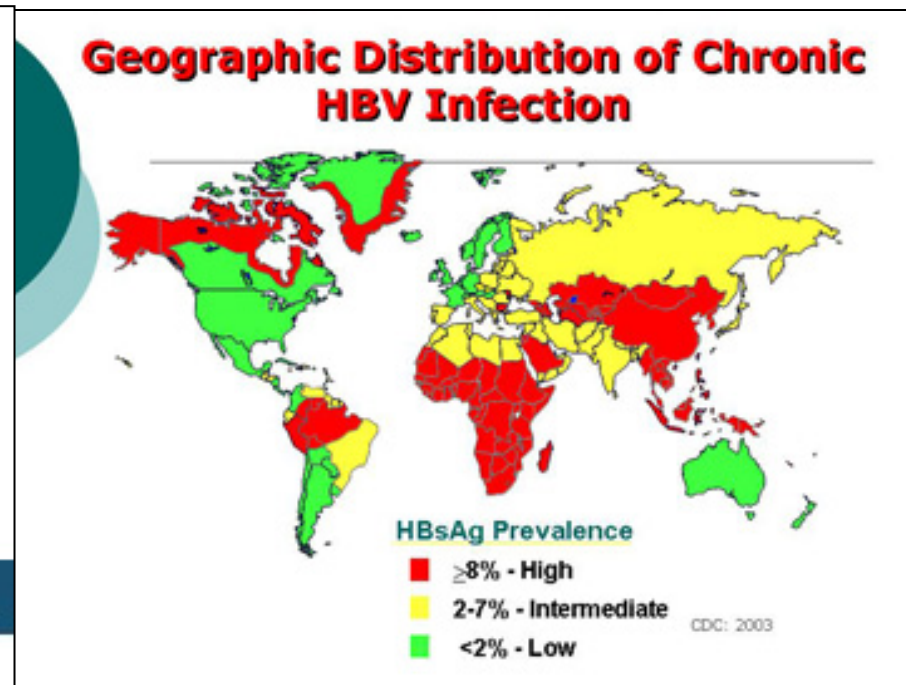
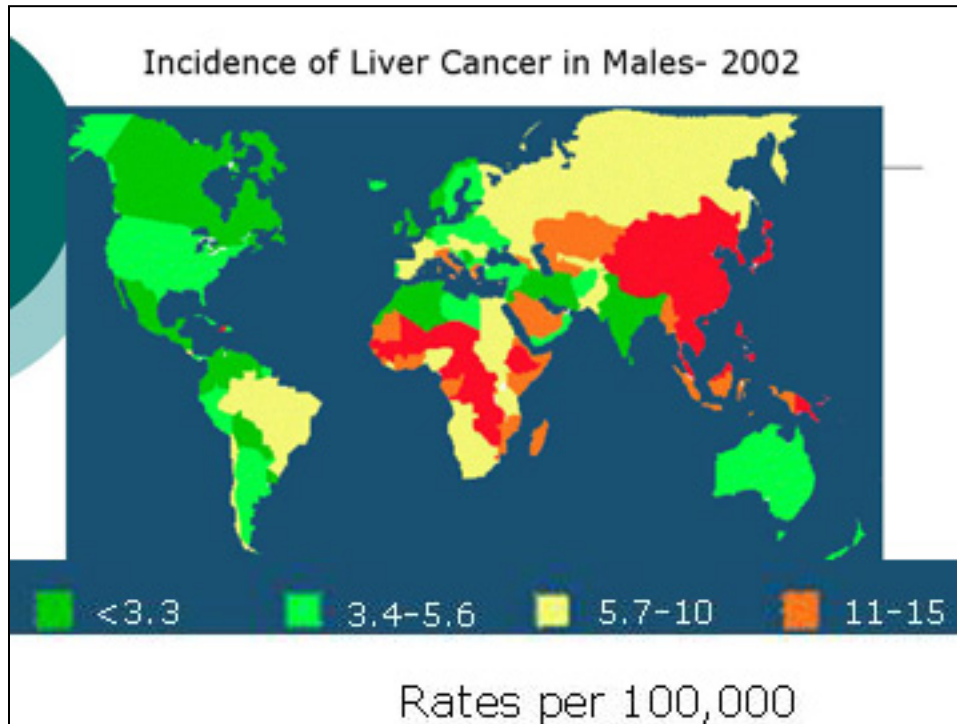
- A. Jemal et al., “Recent Geographic Patterns of Lung Cancer and Mesothelioma Mortality Rates in 49 Shipyard Counties in the U.S., 1970-94”, Am J. Ind. Med. 2000, 37(5):512-21.
- Lung cancer mortality rate among white males elevated during 1950-69
  - in counties with WW2 shipbuilding industry
  - Risk associated with asbestos exposure





# Colocation Example

- Example: Liver Cancer, HBV Infection



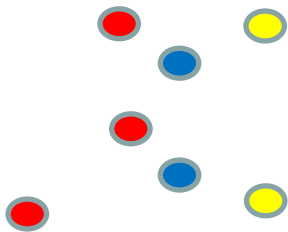


# Examples of Spatial Associations, Colocations

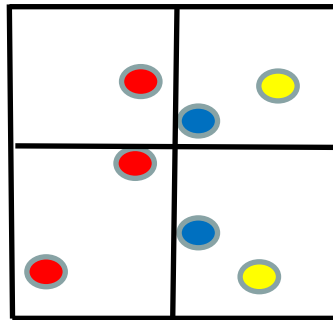
- Spatial Associations, Colocations
  - Tobacco farming, Tobacco use
  - Appalachian OH, high poverty, few healthcare professionals, mammography centers
  - Food deserts, increased rate of obesity & cancer
  - Lower quality of mammography centers in Afr. Am. Chicago neighborhoods
  - Access to state of the art cancer treatment centers limited by distance
- Sources: (a) Place as a risk factor: how Geography shapes where cancer strikes, Elektra Paskett, [www.nyp.org/cancer/cancerprevention/cancer-prevention-articles/029-how-geography-shapes-where-cancer-strikes](http://www.nyp.org/cancer/cancerprevention/cancer-prevention-articles/029-how-geography-shapes-where-cancer-strikes);
- (b) Breaking the cycle of despair One woman's battle for the health of Appalachia, By Bob Tedeschi, June 20, 2016. <https://www.statnews.com/2016/06/20/breaking-cycle-despair-one-womans-battle-health-appalachia/>

# Traditional Statistics is limited for Spatial Interactions!

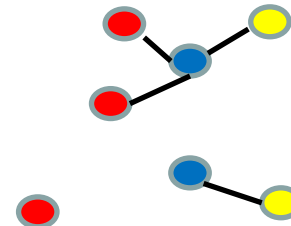
- Challenge: Traditional statistical methods miss spatial interactions
- Prelim. Results: Co-location and teleconnection reveal spatial interaction
  - between variables for point data types
- Proposed: address data with multiple levels of aggregation, e.g., areal summary







(a) a map of 3 features



(b) Spatial Partitions



(c) Neighbor graph

	Pearson's Correlation	Ripley's cross-K	Participation Index (colocation)
 - 			
 - 			

# Fast Algorithms to Mine Colocations from Big Data

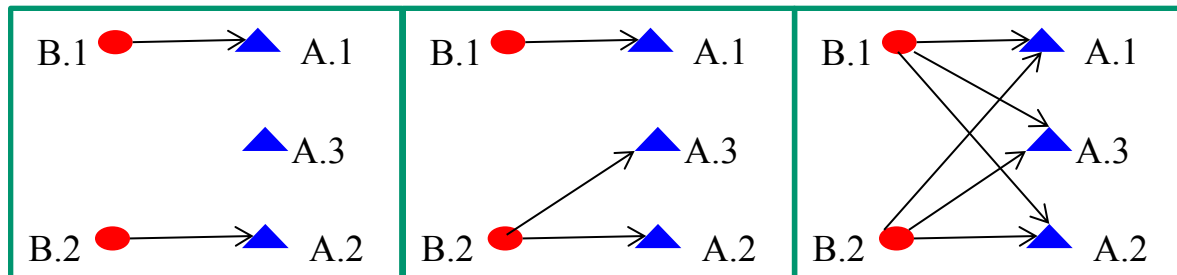
**Participation ratio**  $pr(f_i, c)$  of feature  $f_i$  in colocation  $c = \{f_1, f_2, \dots, f_k\}$ :  
 fraction of instances of  $f_i$  with feature  $\{f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k\}$  nearby

**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

# Outline

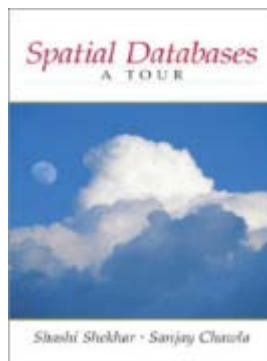
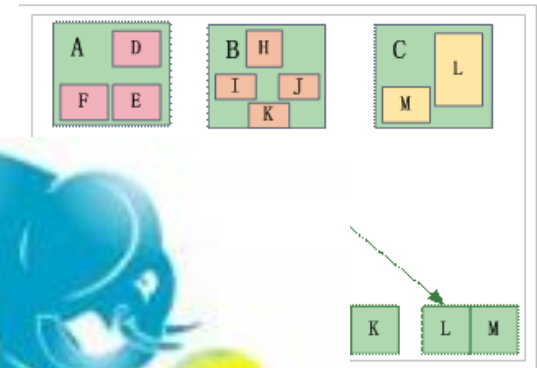
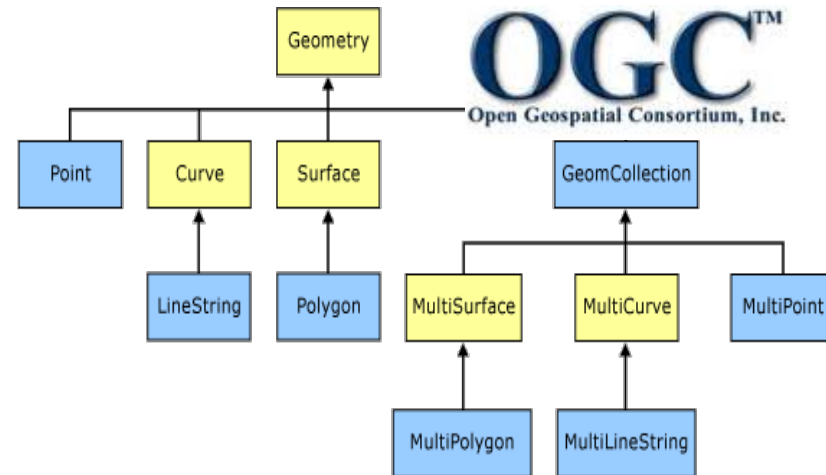


- Introduction
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Statistics
  - Spatial Database Management Systems
    - Scalability => Privacy
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
- Conclusions



# Spatial Database Management

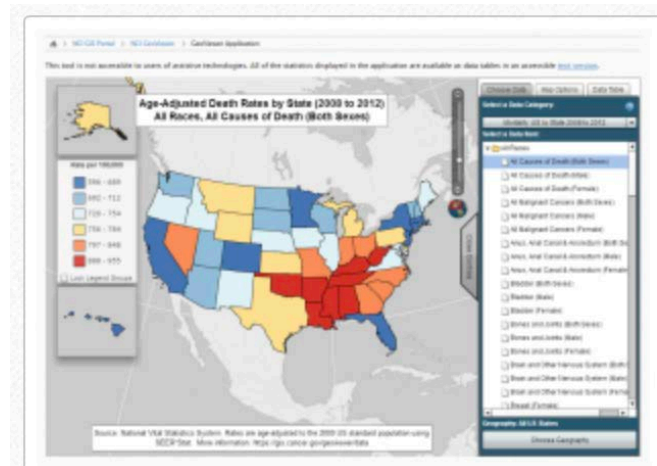
- Meta-data, Schema, DBMS (SQL, Hadoop)
- Challenge: One size does not fit all!
- Ex. Spatial Querying
  - Geo-tag. Checkin, Geo-fence
- Spatial Querying Software
  - OGC Spatial Data Type & Operations
  - Data-structures: B-tree => R-tree
  - Algorithms: Sorting => Geometric
  - Partitioning: random => proximity aware



# Spatial Databases in NCI GeoViewer Tool

- 
- Available Statistics
  - Demographics from the American Community Survey for 2011 to 2015
  - SEER Incidence for 2010 to 2014 for the entire U.S. for available state registries
  - U.S. Mortality for 2010 to 2014 for the entire United States
  - U.S. Historical Mortality for 1970 to 2009 for the entire United States
  - Prevalence from the National Center for Health Statistics for 2017
  - Screening & Risk Factors from BRFSS (CDC Behavioral Risk Factor Surv. System)

- Source: <https://gis.cancer.gov/geoviewer/>



# 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

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)



**GEOTARGETED**  
**ALERTS AND WARNINGS**



# Outline

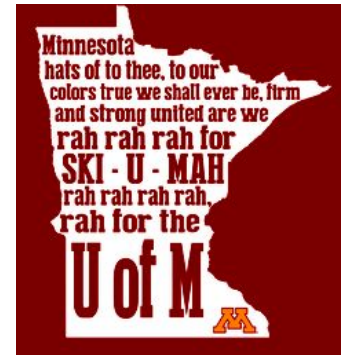


- Introduction
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Statistics
  - Spatial Database Management Systems
  - Virtual Globes & Remote Sensing
    - Quilt => Time-travel & Depth
  - Geographic Information Systems
- Conclusions



# 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, ...



# 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 core hours for global 29-frame video



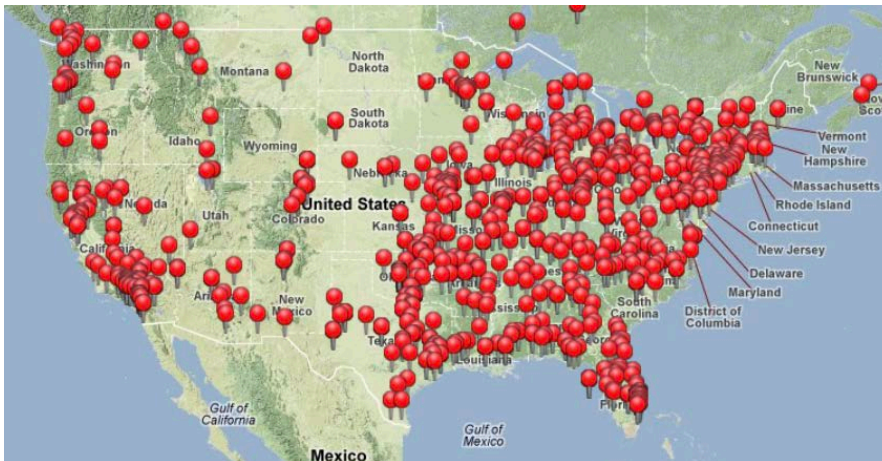
[googleblog.blogspot.com/2013/05/a-picture-of-earth-through-time.html](http://googleblog.blogspot.com/2013/05/a-picture-of-earth-through-time.html)

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



# 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, ...



People's map of reported cancer (Source: Brokovicz.com)

# Virtual Globes in GIS Education

- Coursera MOOC: From GPS and Google Earth to Spatial Computing
  - 21,844 students from 182 countries (Fall 2014)
  - 8 modules, 60 short videos, in-video quizzes, interactive examinations, ...
  - 3 Tracks: curious, concepts, technical
  - Flipped classroom in UMN on-campus course





# Outline

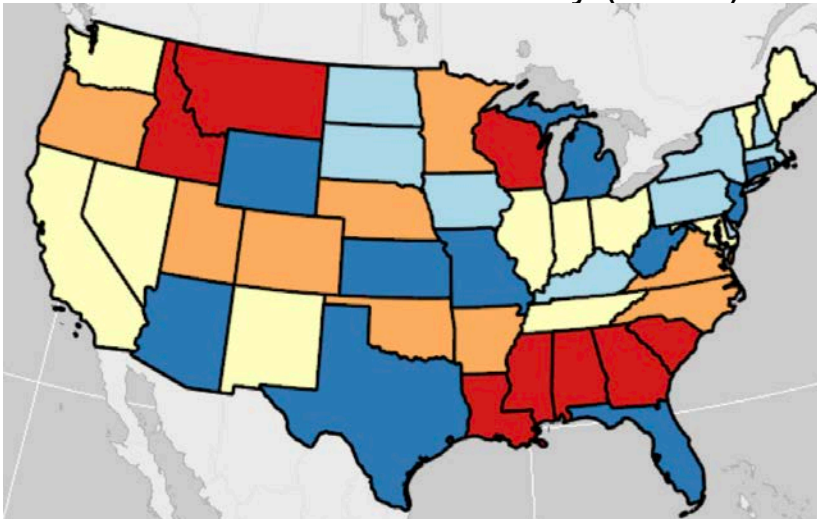


- Introduction
- Broad Interest Examples
  - GPS
  - Location Based Services
  - Spatial Statistics
  - Spatial Database Management Systems
  - Virtual Globes & Remote Sensing
  - Geographic Information Systems
    - Geo => Beyond Geo
- Conclusions

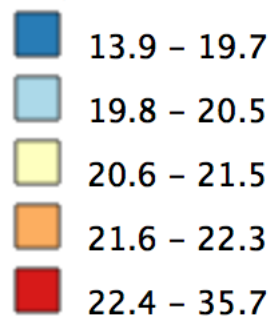
# NCI Cancer Maps – State level

- NCI Geographic Info. Systems & Sciences for Cancer Control
- Maps for lung, breast, prostate, lung cancer, colorectal, and cervical cancers

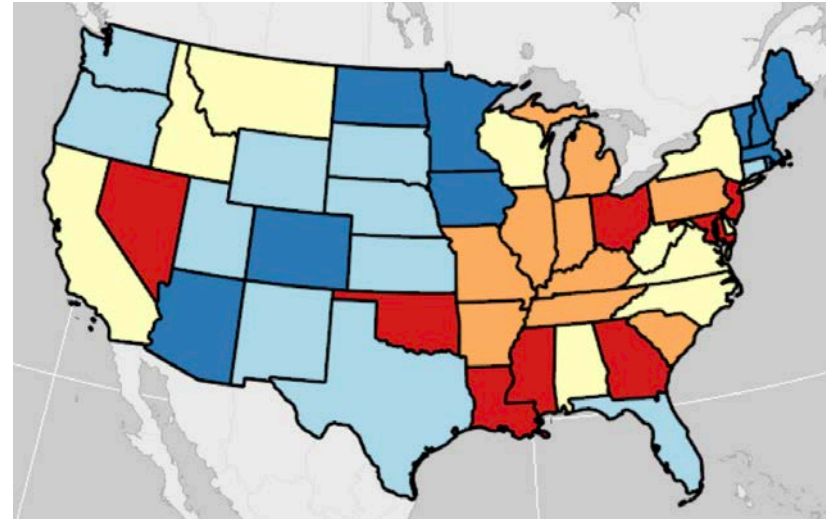
Prostate Cancer Mortality (09-13)



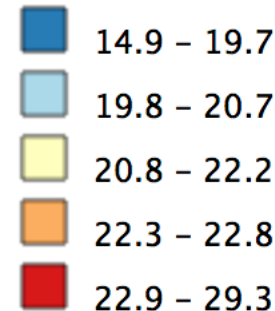
State Age-Adjusted Rate (per 100,000)



Breast Cancer Mortality (09-13)

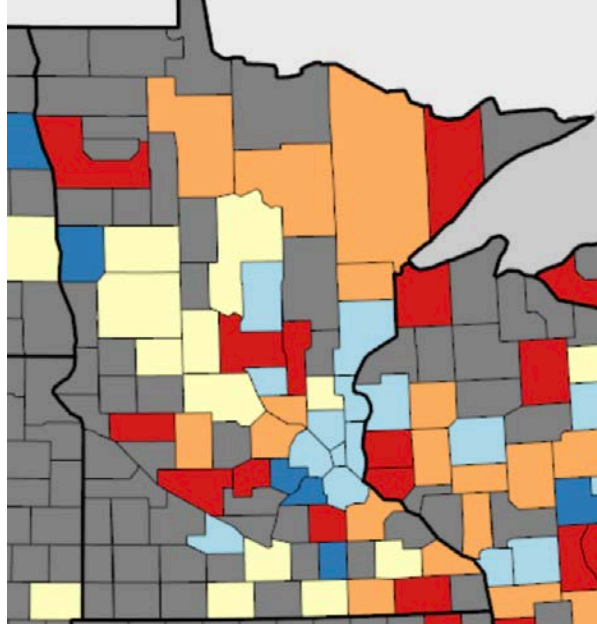


State Age-Adjusted Rate (per 100,000)

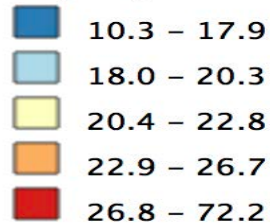


# NCI Cancer Maps – County level

Prostate Cancer Mortality (09-13)



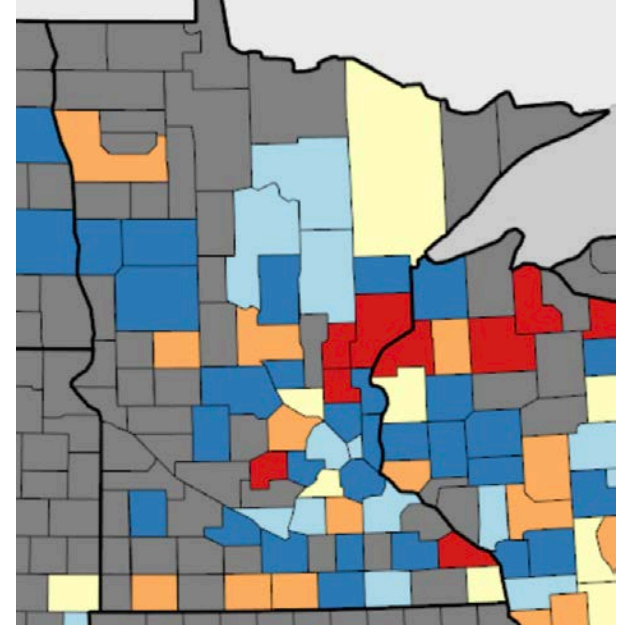
**County Age-Adjusted Rate (per 100,000)**



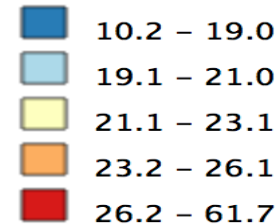
**Suppressed**



Breast Cancer Mortality (09-13)



**County Age-Adjusted Rate (per 100,000)**



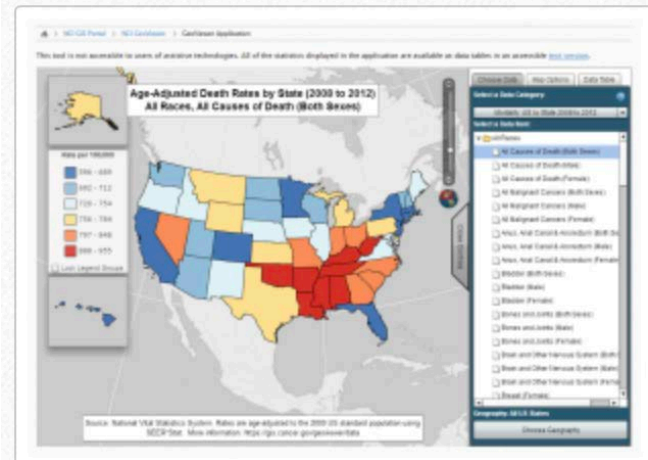
**Suppressed**



- Caution: Suppressed areas may include interesting information.

# NCI GeoViewer Tool

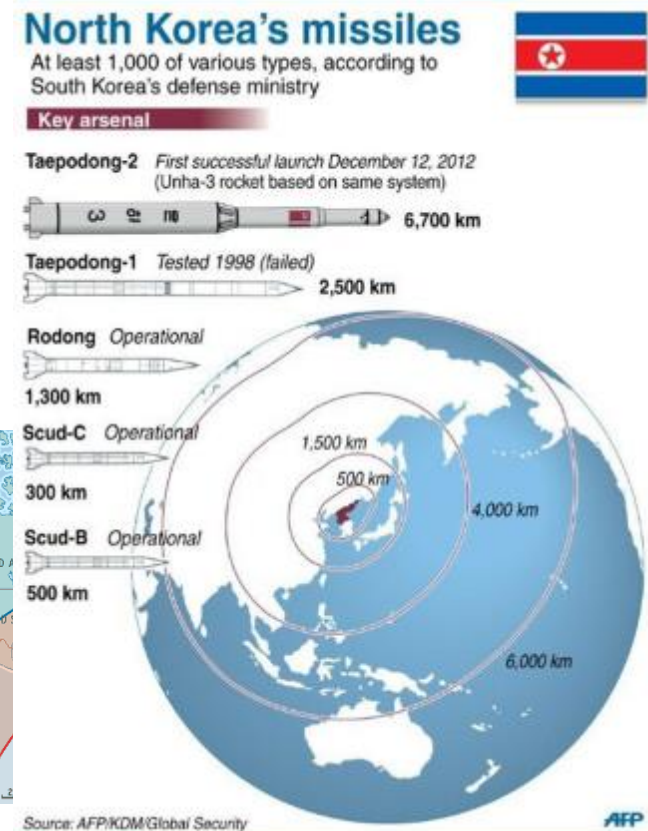
- Creates maps of cancer statistics, demographics, and risk factors.
  - Select the type of Statistic and By-Variables shown on the map
  - Map option to change colors and groups of mapped statistics
  - Extract a map or download statistics



- 
- Available Statistics
  - Demographics from the American Community Survey for 2011 to 2015
  - SEER Incidence for 2010 to 2014 for the entire U.S. for available state registries
  - U.S. Mortality for 2010 to 2014 for the entire United States
  - U.S. Historical Mortality for 1970 to 2009 for the entire United States
  - Prevalence from the National Center for Health Statistics for 2017
  - Screening & Risk Factors from BRFSS (CDC Behavioral Risk Factor Survey System)
- Source: <https://gis.cancer.gov/geoviewer/>

# 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.
  - Q? How to model time? Spatio-temporal?
- **Reference Systems**
  - Which countries in North Korea missile range?
  - 3D Earth surface displayed on 2D plane
  - Spherical coordinates vs. its planar projections
  - Q? What are reference systems for time?



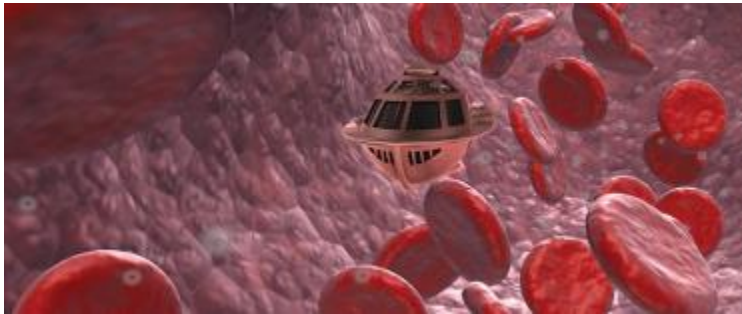
**The Economist**



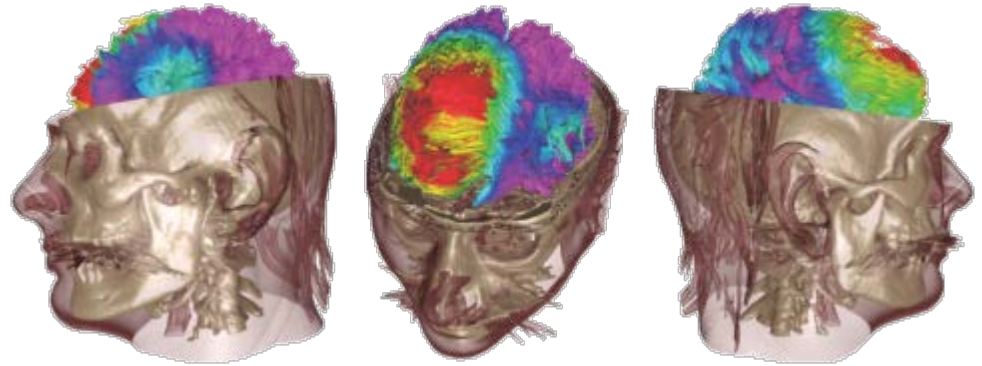
# Opportunities: Beyond Geographic Space

- 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 & Remote Sensing
- Geographic Information Systems
- **Conclusions**

# Place as a risk factor

- A success story – Delaware
  - Eliminated racial disparity in colorectal cancer screening, incidence and death rates
  - Took 10 years and some legislation
- Spatial variations in risky behavior, e.g., Smoking, Obesity, HPV vaccination
- Hotspots of colorectal cancer mortality [Am Cancer Society]
  - Mississippi Delta, Appalachia, and VA/NC
- Spatial Associations
  - Tobacco farming, Tobacco use
  - Appalachian OH, high poverty, few healthcare professionals, mammography centers
  - Food deserts, increased rate of obesity & cancer
  - Lower quality of mammography centers in Afr. Am. Chicago neighborhoods
  - Access to state of the art cancer treatment centers limited by distance
- Sources: (a) Place as a risk factor: how Geography shapes where cancer strikes, Elektra Paskett, [www.nyp.org/cancer/cancerprevention/cancer-prevention-articles/029-how-geography-shapes-where-cancer-strikes](http://www.nyp.org/cancer/cancerprevention/cancer-prevention-articles/029-how-geography-shapes-where-cancer-strikes);
- (b) Breaking the cycle of despairOne woman's battle for the health of Appalachia, By Bob Tedeschi, June 20, 2016. <https://www.statnews.com/2016/06/20/breaking-cycle-despair-one-womans-battle-health-appalachia/>

# Spatial Context Matters



**Francis S. Collins** ✓

@NIHDirector

 Follow

T. Glass: If DNA is our biological blueprint, ZNA (zipcode at birth) is the blueprint for behavioral&psycho-social makeup.

[#PMINetwork](#)

10:10 AM - May 29, 2015

**National Institutes of Health**

**Funding Opportunity**

**Spatial Uncertainty: Data, Modeling, and Communication**

**(Sponsoring Institutes: NCI, NHLBI, NIAAA, NIAID, NICHD, NIDA, and NIEHS)**

PA 15-09 (R21), PA 15-10 (R01), PA 15-011 (R03),

**NIH BD2K (Big Data to Knowledge) Initiative**

<https://datascience.nih.gov/bd2k>

# (Spatial)-Data Driven Hypothesis Generation

1854: What causes Cholera?

? water pump

Remove pump handle



Collect & Curate Data



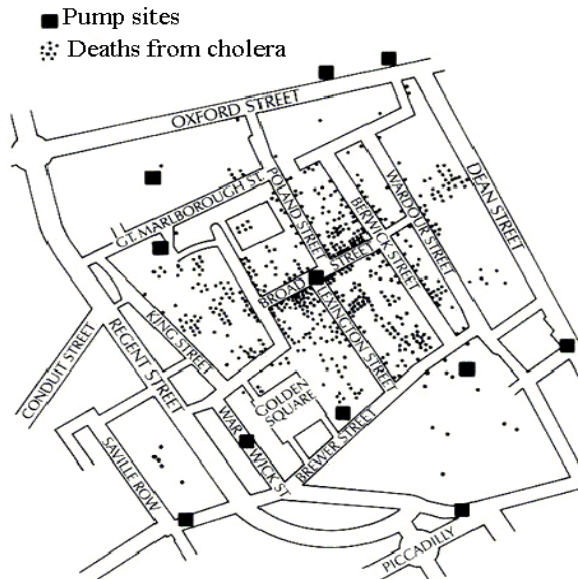
Discover Patterns, Generate Hypothesis



Test Hypothesis (Controlled Experiments)



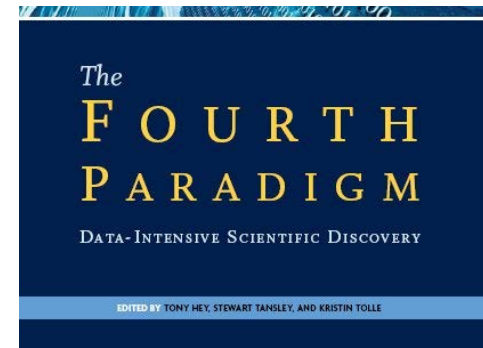
Develop Theory



Today: What are side-effects of a treatment?

NIH BD2K (Big Data to Knowledge) Initiative

<https://datascience.nih.gov/bd2k>





# Primary Sources

1. Spatial Computing, Communications of the ACM, 59(1), Jan. 2016.
2. Identifying patterns in spatial information: a survey of methods ([pdf](#)), [Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery](#), 193-214, 1(3), May/June 2011. (DOI: 10.1002/widm.25).
3. From GPS and Virtual Globes to Spatial Computing 2020, Computing Community Consortium Report, 2013. [www.cra.org/ccc/visioning/visioning-activities/spatial-computing](http://www.cra.org/ccc/visioning/visioning-activities/spatial-computing)

