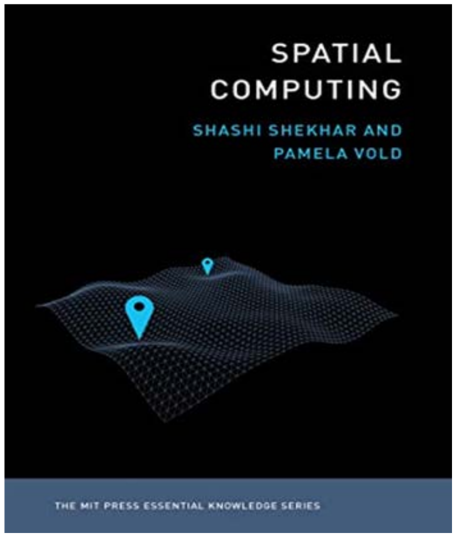
University of Chicago - Research Computing Center Speaker Series
(Theme: *Computing, Data and Beyond- Impact on Our World)*
Feb. 21st, 2022



What's Special About GeoAI and Spatial Data Science?

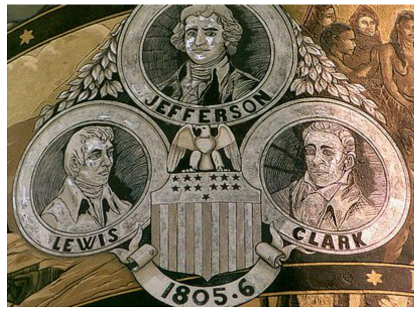
Shashi Shekhar
McKnight Distinguished University Professor, Univ. of Minnesota
www.cs.umn.edu/~shekhar

University of Minnesota
Driven to Discover®

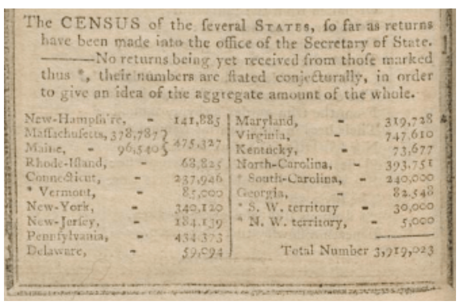
Acks.: NSF, USDOD-NGA, USDOE-ARPA-E, USDA-NIFA, NASA, …

Happy President's Day!

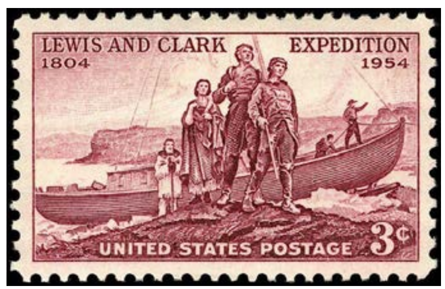
We are grateful to many great presidents for transformative contributions such as Spatial Data and Geo-Intelligence



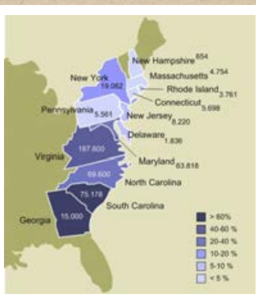
… Eisenhower … Clinton …

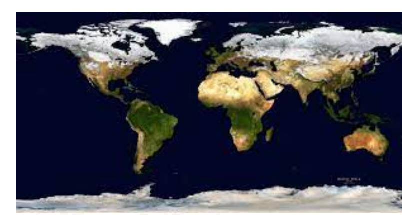

The CENSUS of the ſeveral STATES, ſo far as returns have been made into the office of the Secretary of State. ——No returns being yet received from thoſe marked thus *, their numbers are ſtated conjecturally, in order to give an idea of the aggregate amount of the whole.

New-Hampſhire,	141,885	Maryland,	319,728
Maſſachuſetts, 378,787 / Maine, 96,540	475,327	Virginia,	747,610
		Kentucky,	73,677
Rhode-Iſland,	68,825	North-Carolina,	393,751
Connecticut,	237,946	* South-Carolina,	240,000
* Vermont,	85,000	Georgia,	82,548
New-York,	340,120	* S. W. territory	30,000
New-Jerſey,	184,139	* N. W. territory,	5,000
Pennſylvania,	434,373		
Delaware,	59,094	Total Number	3,919,023



1960: Eisenhower reviews photo from Satellite Tiros I.



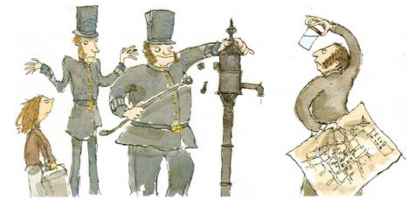


Acknowledgements

- P.I., Connecting the Smart-City Paradigm with a Sustainable Urban Infrastructure Systems Framework to Advance Equity in Communities, NSF (1737633), $2.5 M, 9/1/2017 - 8/31/2021.
- P.I., Spatio-temporal Informatics for Transportation Science, NSF (1901099), $1.2M, 8/1/19-7/31/23.
- P.I., EAGER: Spatiotemporal Big Data Analysis to Understand COVID-19 Effects, $100K, NSF (2040459), 9/1/20-8/31/22.
- P.I., Identifying Aberration Patterns in Multi-attribute Trajectory Data with Gaps, $600K, USDOD-NGA (HM0476-20-1-0009), 6/15/20- 6/14/23.
- Co-P.I., HDR Institute: iHARP- Harnessing Data and Model Revolution in the Polar Regions, $13M, NSF (2118285), 9/15/2021 – 8/31/2026.
- Co-P.I., WinterTurf: A Holistic Approach to Understanding the Mechanisms and Mitigating the Effects of Winter Stress on Turfgrasses in Northern Climate, $8M, NIFA (2021-51181-35861), 9/1/21-8/31/26.

A **Geo-Intelligence** and **Spatial Data Science** Story

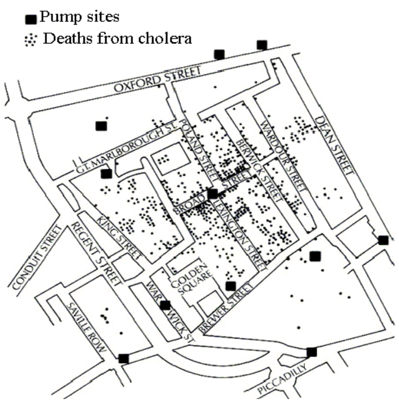
1854: What causes Cholera?

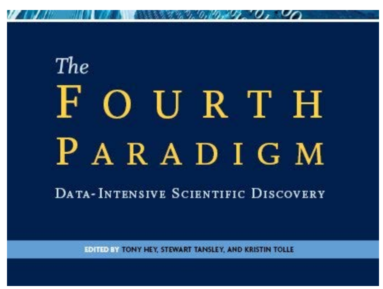


Collect & Curate Data → Discover Patterns, Generate Hypothesis → Test Hypothesis (Experiments) → Develop Theory

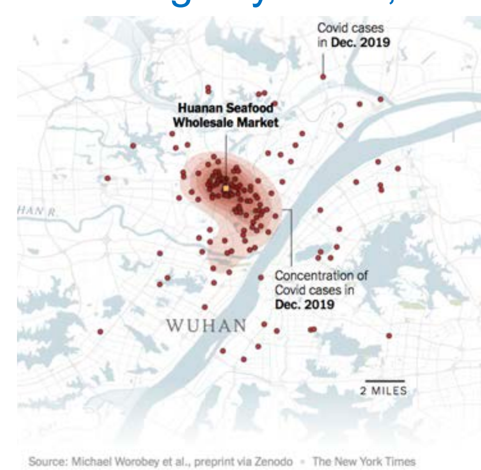
? water pump

Remove pump handle





Impacts: Hygiene, Separate drinking water and sewage systems, …



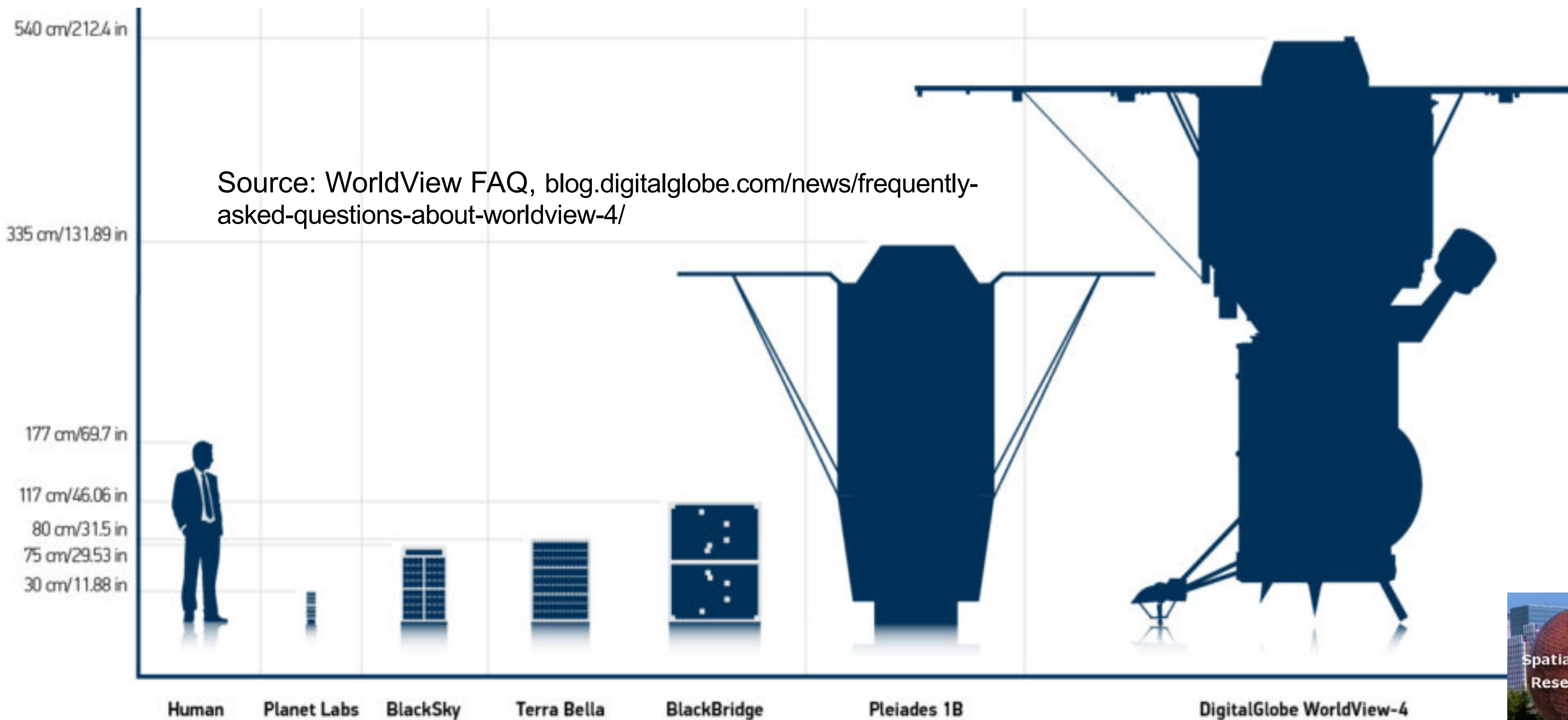
Q? What are Choleras of today?
Q? How may Geo-Intelligence and Spatial Data Sc. Help?

What has changed?

	Before	Now
Spatial Data Revolution	Smaller data from surveys, few satellites and sensors	**Spatial Big Data** from Nano-satellites, Billions of GPS enabled devices, …
Better and Pl		
Spatial Proces		
Spatial Data Science		
Spatial Data Visualization		

Large Constellations of Small Satellites

- Hi-frequency (e.g., daily or hourly) time-series of imagery of entire earth
- Small Satellites: video (5-minutes): https://geospatialstream.com/sciencecasts-nasa-embraces-small-satellites/
- Large Constellations
  - 2021: Planet Labs: 200+ satellites: daily Earth scan (1m resolution, visible+NIR bands)



Source: WorldView FAQ, blog.digitalglobe.com/news/frequently-asked-questions-about-worldview-4/

Spatial Computing Research Group

GPS is Ubiquitous 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

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*

Spatial Data Revolution

- Remotely sensed Imagery
  - Thousands of (Nano-)satellites
  - UAVs, Aerial imagery, …
- (GPS-) Location traces
  - Billions of phones, vehicles, …
  - Spatio-temporal patterns of life
- Others
  - Vehicle On-board diagnostics
  - Geo-social media, …
- Why is it interesting?
  - See previously inconspicuous
  - Monitor hard to monitor areas
  - Solve previously unsolvable problems

The New York Times

Published: May 13, 2011

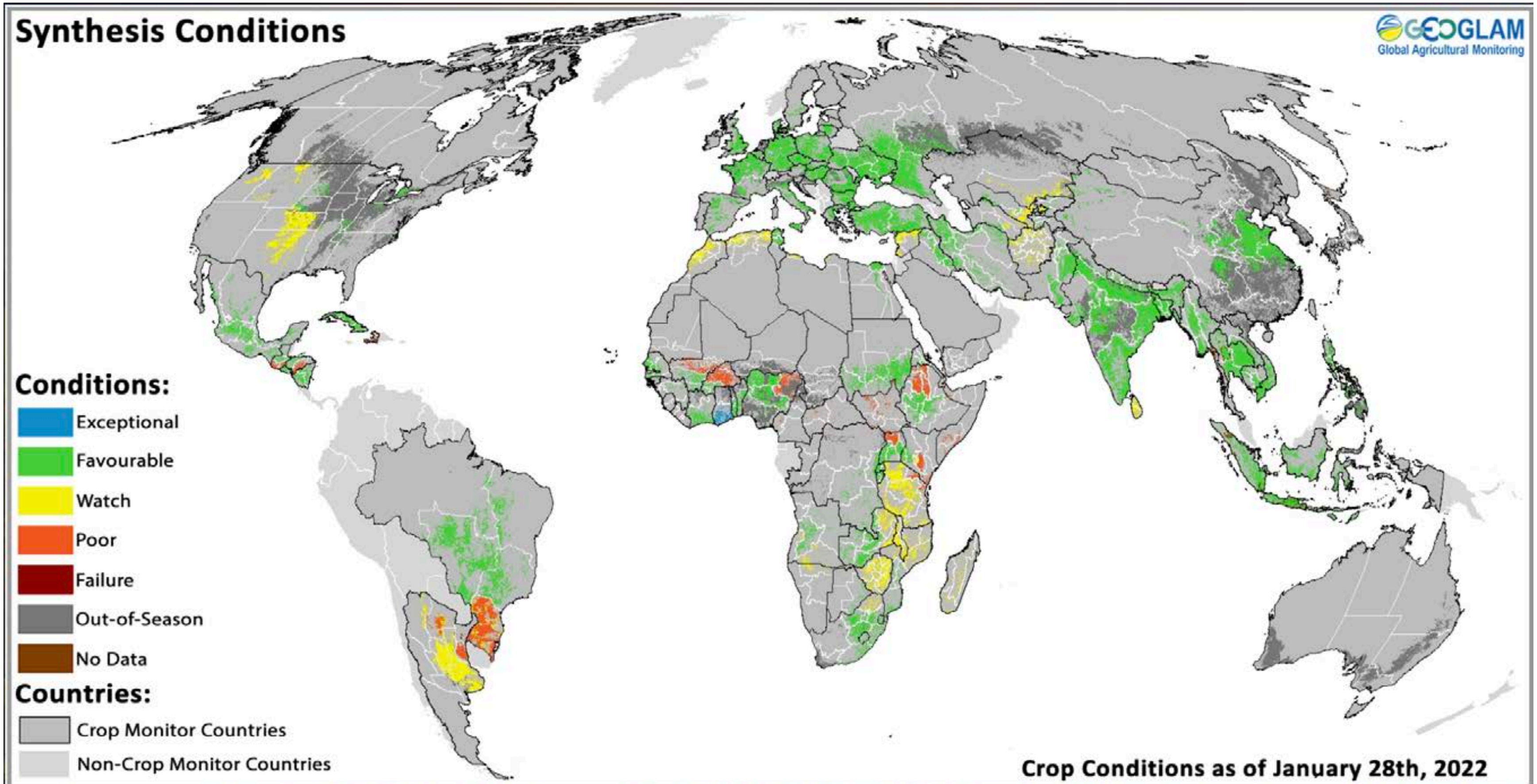
McKinsey Global Institute

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.

Source: Y. Xie et al., [Transforming Smart Cities With Spatial Computing](), Proc. [IEEE Intl. Conf. on Smart Cities](), 2018.

Monitor Global Crops for Early Warning

- Last century: US Wetland inventory took 4 decades and $400M
- Now: Global crop-health maps produced monthly for early warning and action

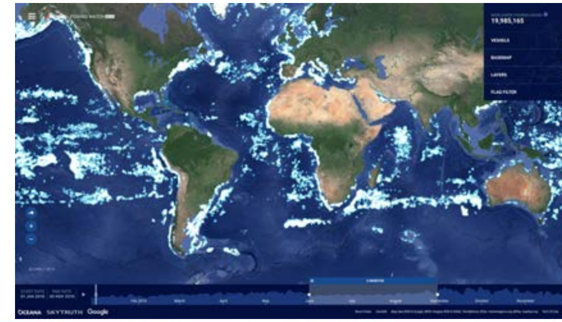


AMIS Agricultural Market Information System

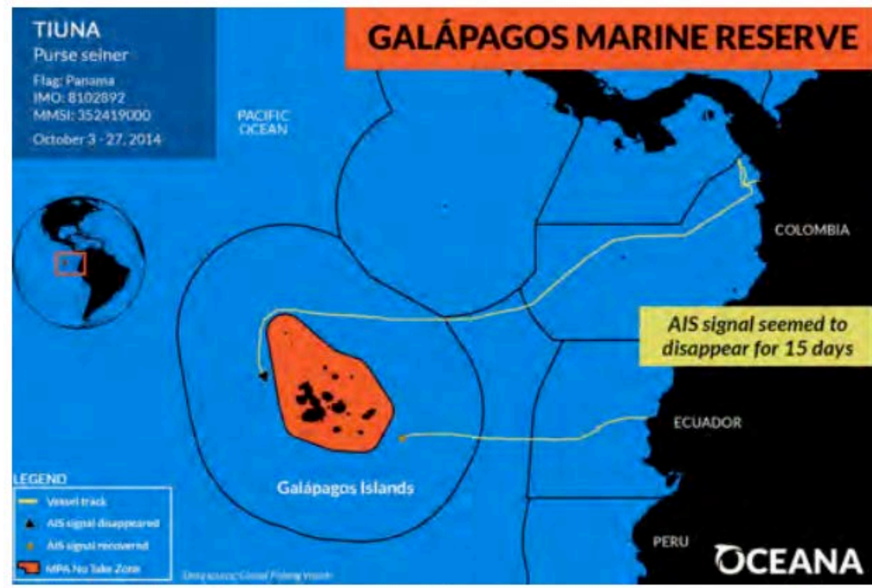
GEOGLAM Global Agricultural Monitoring

Better Visibility of Activities at Oceans and Seas

- "For years it's been impossible to see illegal acts happening at sea, from overfishing to human rights abuses. Now that's changing" (Source (b))



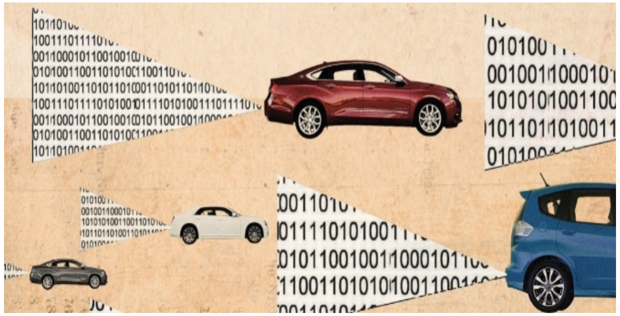
- **Automatic Identification System (AIS):**
  - Ships (> 300 tons) report location
  - Collision Avoidance (augment marine radar)
  - Monitor fishing and cargo fleet
  - Search and Rescue, Statistics and Economics



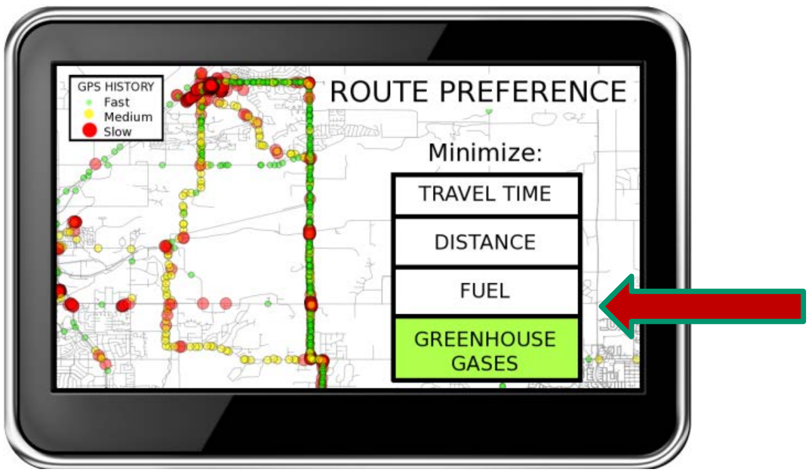
- **Example**:
  - A fishing vessel switched off AIS for 15 days.
  - near the Galapagos Marine Reserve

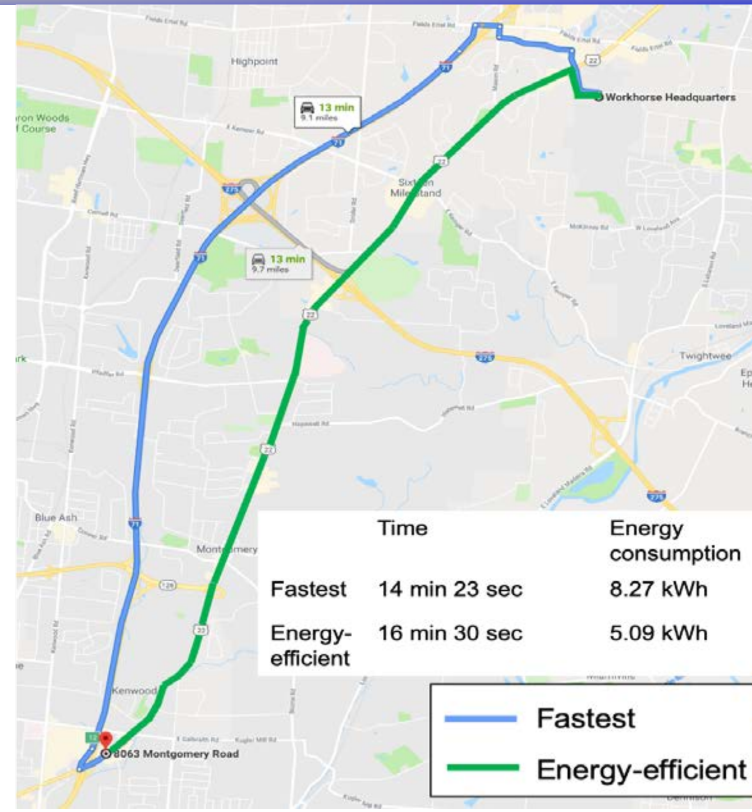
Sources: *(a)* [How Illegal Fishing Is Being Tracked From Space](), Natl. Geographic, 3/12/2018.
(b) [How to spot the secretive activities of rogue fishing boats](), bbc.com, 7 June 2018,
(c) [Intelligence Agencies Pushed to Use More Commercial Satellites](), New York Times, 9/27/2021.

Monitor energy use and emissions => Eco-Routing



GPS Tracks + On Board Diagnostics Data



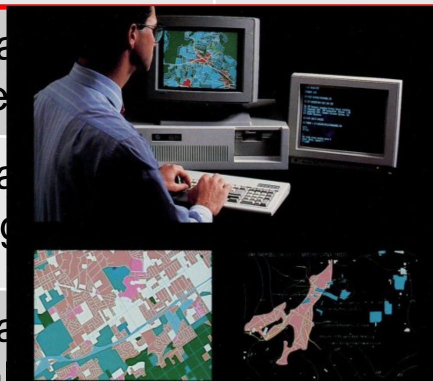


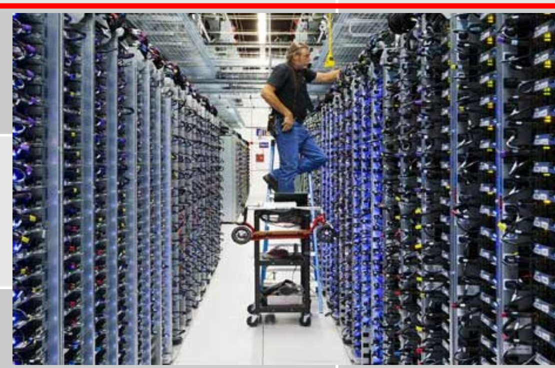
- Oct. 2021: Google Maps supports Eco-routes

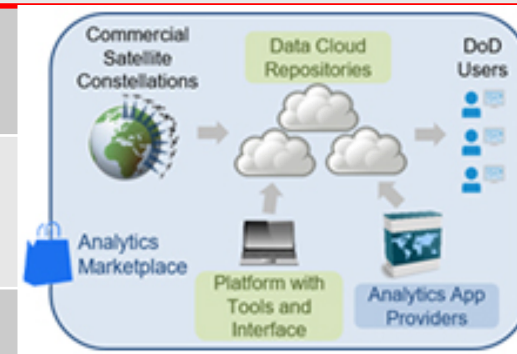
Details: Y. Li, P. Kotwal, P. Wang, Y. Xie, S. Shekhar, and W. Northrop, [Physics-guided Energy-efficient Path Selection Using On-board Diagnostics Data], ACM/IMS Trans. Data Sc. 1(3):1-28, Article 22, Oct. 2020. (Initial results appeared in Proc. ACM SIG-Spatial, 2018).

What has changed?

	Before	Now
Spatial Data Revolution	Smaller data	Spatial Big Data
Better Access and Platforms	Smaller platforms, e.g., ESRI Arc/Info, Postgis	Big Compute, e.g., AWS Earth, ESRI GIS Tools for Hadoop,
Spatia Proce		
Spatia Minin		
Spatia Visual		

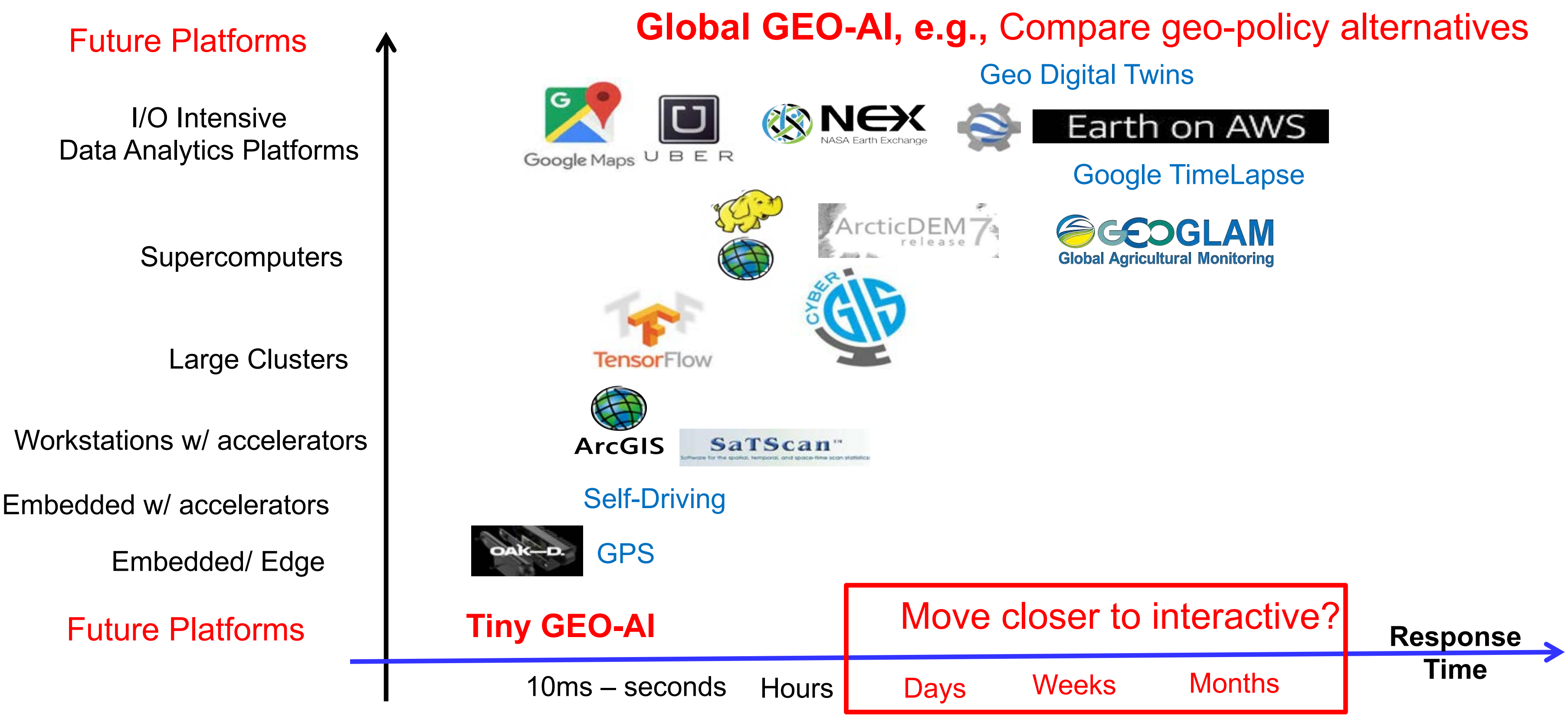






Source: Original PC ARC/INFO Brochure, OCT. 22, 2010 / M. ARTZ, HTTPS://GISANDSCIENCE.COM/2010/10/22/ORIGINAL-PC-ARCINFO-BROCHURE/

Geo-AI Platforms



Note: Accelerators include GPU, TPU, FPGA

Cloud Repositories expand Access to Spatial Big Data

- 2008: USGS gave away 35-year LandSat satellite imagery archive
  - Analog of public availability of GPS signal in late 1980s
- 2017: Cloud-based repositories of geospatial data
  - Explosion in machine learning on satellite imagery to map crops, water, buildings, roads, …

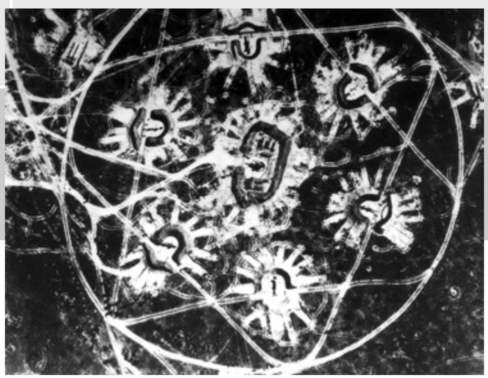
	Google Earth Engines	NEX	AWS Earth
Elevation, Landsat, LOCA, MODIS, NAIP	x	x	x
NOAA	x		x
AVHRR, FIA, GIMMM, GlobCover, NARR, TRIMM, Sentinel-1	x	x	
IARPA, GDELT, MOGREPS, OpenStreetMap, Sentinel-2, SpaceNet (building/road labels for ML)			x
CHIRPS, GeoScience Australia, GSMap, NASS, Oxford Map, PSDI,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

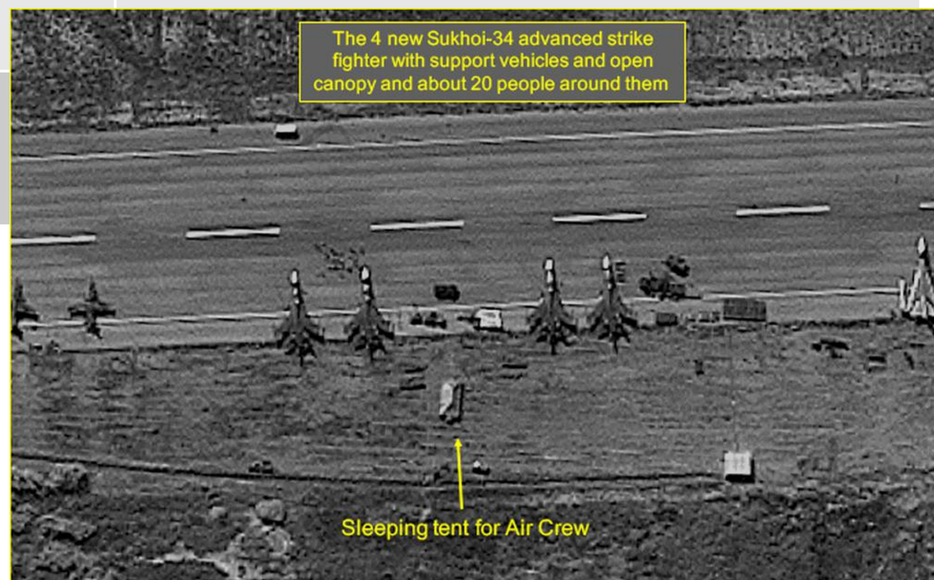
Google Earth Engine

NEX NASA Earth Exchange

What has changed?

	Before	Now
Spatial Data Revolution	Smaller Data	Spatial Big Data
Spatial Data Access and Platforms	Smaller platforms	Big Compute, e.g., Cloud (e.g., AWS Earth, ESRI GIS Tools for Hadoop)
Spatial Data Processing	Fairly manual, labor-intensive (Geo-Intelligence)	More automation (Geo-Augmented-Intelligence (Geo-AI))
Spatial Data Science		
Spatial Data Visualization		





A Geo-AI Development: Object Detection by Deep Learning

NATIONAL ARCHIVES

FEDERAL REGISTER

Federal Register Home

The Daily Journal of the United States Government

Addition of Software Specially Designed To Automate the Analysis of Geospatial Imagery to the **Export Control** Classification Number 0Y521 Series, 01/06/2020, 85 FR 459

Excerpt:

Geospatial imagery "software" "specially designed" for training a *Deep Convolutional Neural Network* to automate the analysis of geospatial imagery and point clouds, and having all of the following:

1. Provides a graphical user interface that enables the user to identify objects (*e.g.,* vehicles, houses, etc.) from within geospatial imagery and point clouds in order to extract positive and negative samples of an object of interest;
2. Reduces pixel variation by performing scale, color, and rotational normalization on the positive samples;
3. Trains a *Deep Convolutional Neural Network* to detect the object of interest from the positive and negative samples; and
4. Identifies objects in geospatial imagery using the *trained Deep Convolutional Neural* Network by matching the rotational pattern from the positive samples with the rotational pattern of objects in the geospatial imagery.

Traditional AI vs. Deep Learning

- From Satellite Imagery: Classify Land-cover, Map buildings
- Ex. 2009 Haiti Earthquake: Map building damage [1]

.

eCognition

Traditional AI

1. Pre-processing
2. Feature engineering & Selection
3. Prediction (class or object)
4. Post-processing (rules, manual)

Deep Learning

1. Pre-processing

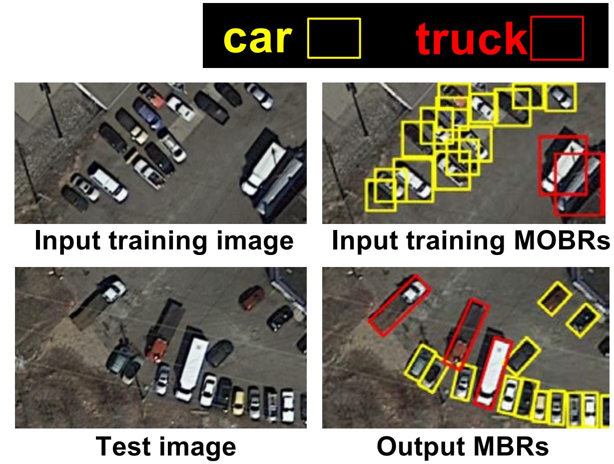
Deep (neural network) Learning

4. Post-processing (guardrail for AI)

[1] J. Aardt et al., Geospatial Disaster Response during the Haiti Earthquake: A Case Study Spanning Airborne Deployment, Data Collection, Transfer, Processing, and Dissemination, Photo. Eng. & Remote Sens., 77(9):943-952, Sept. 2011.

Deep Learning for Geo-Object Detection

- **Q:?** How many vehicles in a parking lot? City?
- **Ex.:** Estimate truck supply in a city (CH Robinson).
- **Old Computer Vision workflow**
  - Many steps, each adds error
- **New Deep Learning Workflow – fewer steps**
  - Aerial imagery (3 inch pixels, Twincites, MN, USA)
  - NAIP Imagery (1 meter pixels, 2017)
    - MA Buildings Data (https://www.cs.toronto.edu/~vmnih/data/)
- Detected Geo-objects
  - Cars, trucks, Houses, …
  - Method: Convolutional Neural Networks (YOLO)



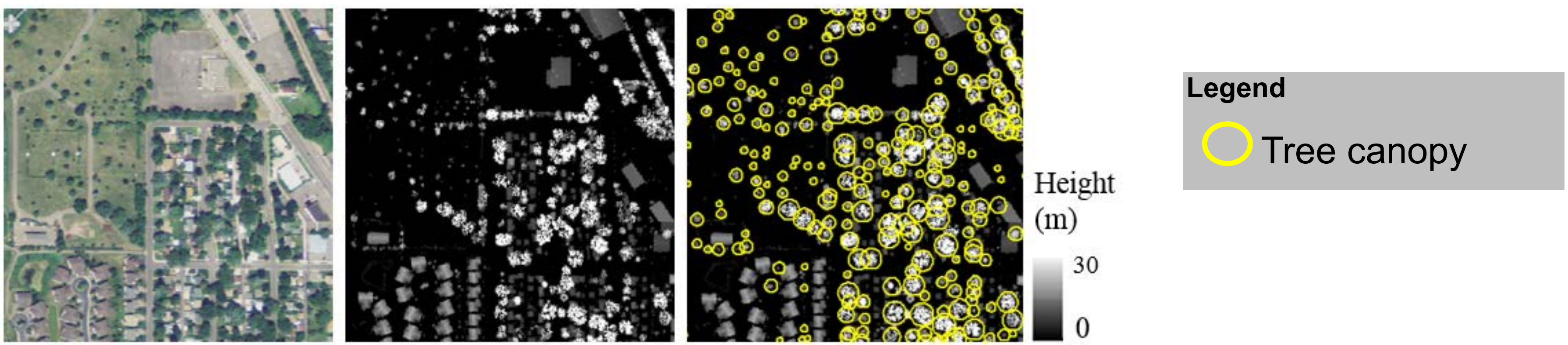
YOLO (baseline) Proposed method

Note: NAIP = National Agriculture Imaging Program (USDA)

Details: Y. Xie et al., An Unsupervised Augmentation Framework for Deep Learning based Geospatial Object Detection: A Summary of Results, Proc. 26th ACM SIGSPATIAL Intl. Conf. Adv. in GIS, 2019.

Mapping Trees from Remote Sensing Imagery

- Why?: Protect Powerlines, Manage Emerald Ash Borer, Green infrastructure Equity
- Input: LiDAR + Remotely Sensed Imagery + (NAIP Ground Truth)
- Approach: Tree Inference by Minimizing Bound-and-band Errors (TIMBER)
  - Optimization to find tree locations and sizes
  - Deep learning constructs features separating trees & non-trees (e.g., light pole)



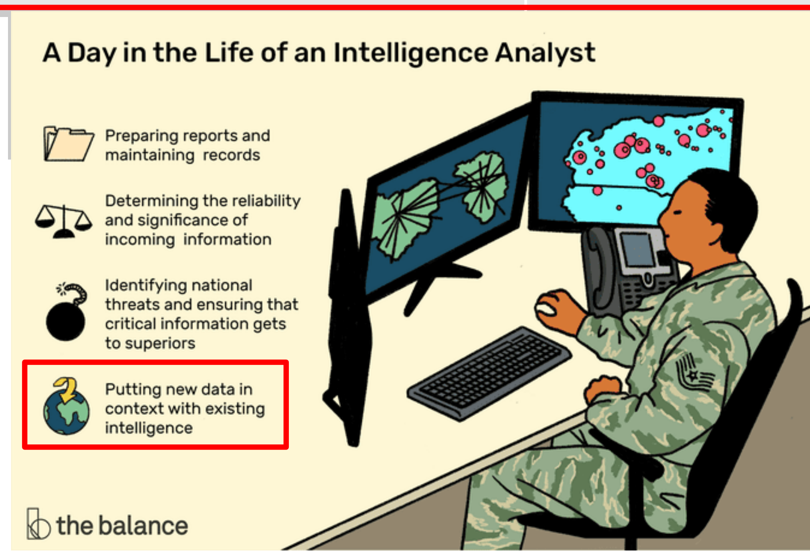
Details: **(a)** Revolutionizing Tree Management via Intelligent Spatial Techniques, Proc. 27th ACM SIGSPATIAL Intl. Conf. on Adv. in GIS, Nov. 2019 Pages 71–74 (Best Vision Paper).

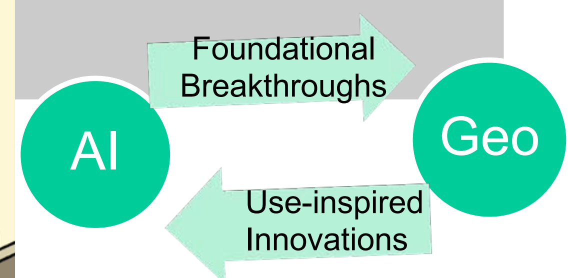
(b) TIMBER: A Framework for Mining Inventories of Individual Trees in Urban Environments using Remote Sensing Datasets, IEEE Intl. Conf. on Data Mining (ICDM), 2018..

What has changed?

	Last Century	This Century
Spatial Data Revolution	Smaller Data	Spatial Big Data
Spatial Data Access and Platforms	Smaller platforms	Big Compute, e.g., Cloud (e.g., AWS Earth, ESRI GIS Tools for Hadoop)
Spatial Data Processing	Fairly manual, labor-intensive (Geo-Intelligence)	More automation Geo-Augmented-Intelligence (Geo-AI)
Spatial Data Science	“One-size fit all” AI applied to spatial data	Virtuous cycle between Geo and AI

AI:



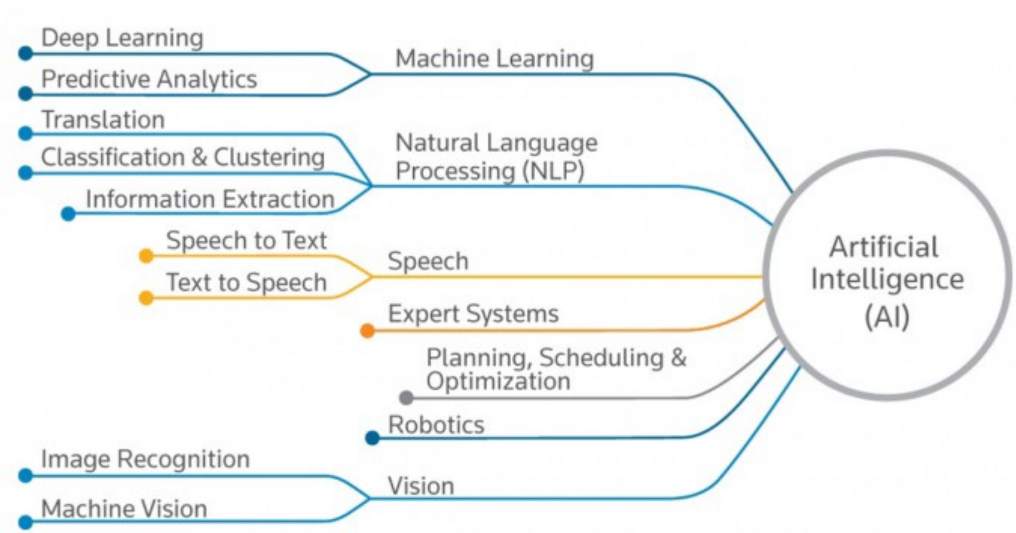


Why go beyond Object Detection?

Source: A. Karpathy (2012), The state of Computer Vision and AI: we are really, really far away
http://karpathy.github.io/2012/10/22/state-of-computer-vision/

Geo-AI beyond Computer Vision

- Research Initiatives
  - USDOE ORNL Trillion Pixel Challenge ; Map all buildings in US and beyond
  - Self-driving: Detect road objects in adverse weather (e.g., rain, snow, dust)
  - 2017-20: DARPA Geospatial Cloud Analytics: Crop yield, fracking, illegal fishing
  - 2020-onwards: IARPA Space-based Machine Automated Recognition Technique (SMART)
    - Change (Construction Stage); Underground: (Subterranean Challenge)
- Text, multimedia, knowledge graphs
  - Question Answering Systems: Answer geographic questions,
  - Geo-locate a picture/video, Fact-check maps, …
- Points, Polygons, Trajectories
  - Characterize spatio-temporal patterns of life



Need for AI (Vision) to break out of "RGB+Lidar" box

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.

The New York Times

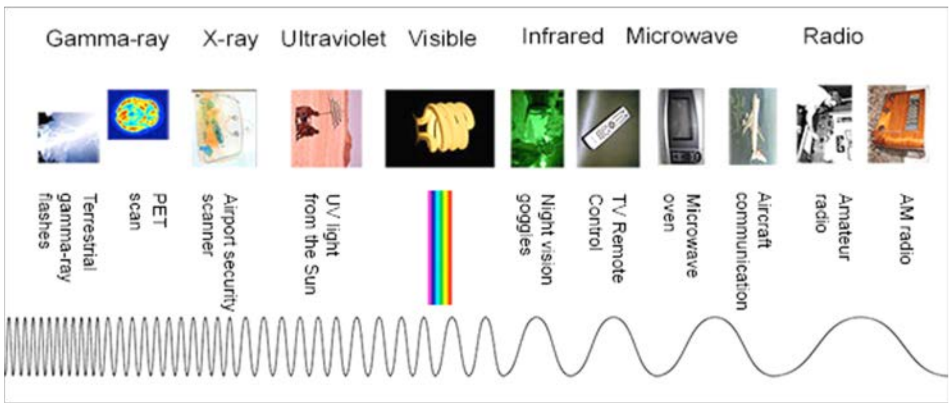
Sept. 26, 2019

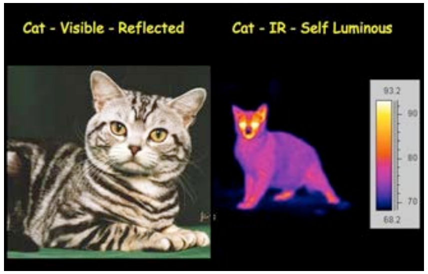
These High-Tech Sensors May Be the Key to Autonomous Cars

Human vision → Superhuman Vision

- Remote sensing use richer Sensors than AI (Computer Vision):
- Ex. Electromagnetic radiation Emitted by objects above absolute zero (0° Kelvin(K), −273°C)

Human





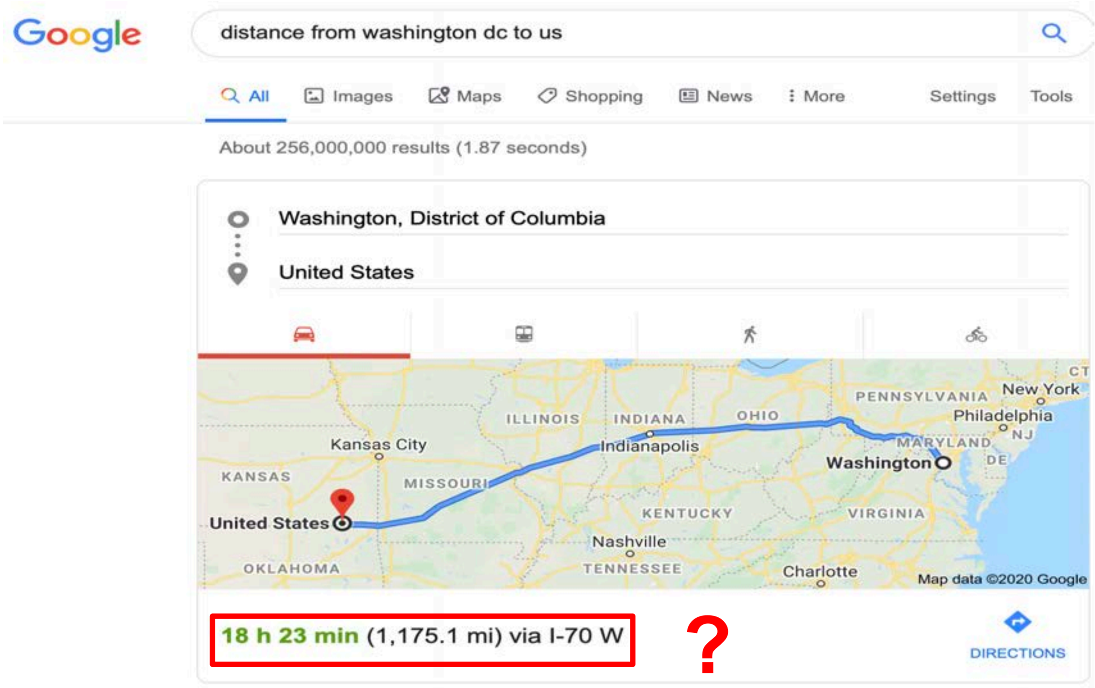
Source: directthermography.co.uk

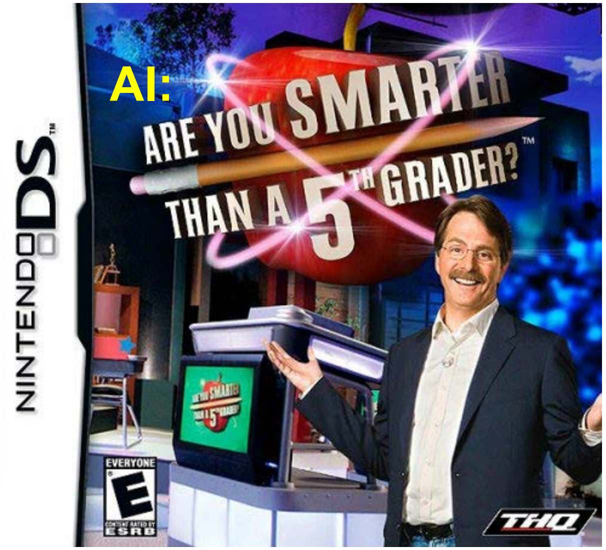
Source: imagine.gsfc.nasa.gov

Should AI learn Richer Spatial Concepts?

Q? What is distance between Washington D.C. and U.S.A.?

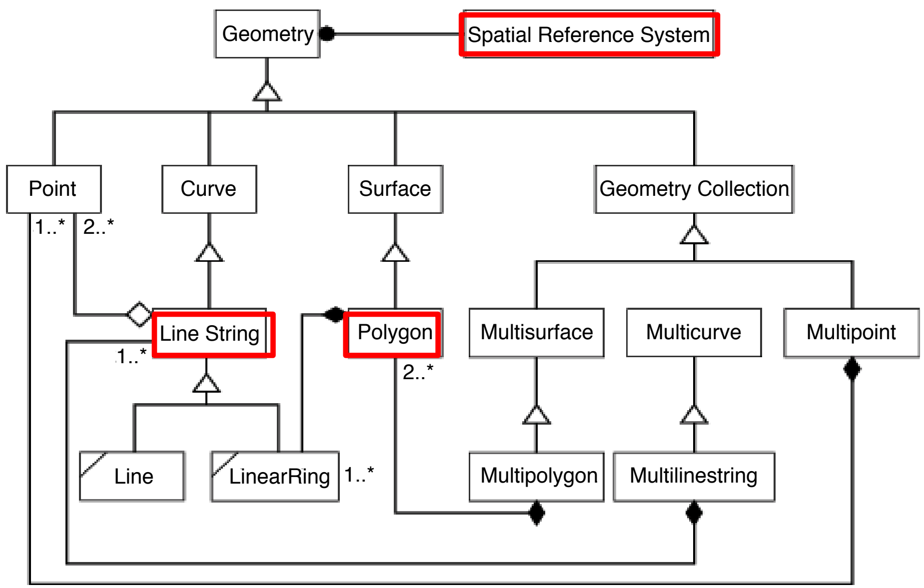
- Zero (Washington D.C. is inside U.S.A.)
- NSF Open Knowledge Networks initiative grants on geo-knowledge networks!





Spatial Data Types: OGC Simple Features Standard

- Spatial Concepts: Point, LineString, Polygon, Collections
- Relationships: Topological, Metric, …
- Helps feature selection for machine learning
  - Ex. Distance to key geo-features, Neighbor relationship



Basic Functions	SpatialReference ()
	Envelop ()
	Export ()
	IsEmpty ()
	IsSimple ()
	Boundary ()
Topological / Set Operators	Equal
	Disjoint
	Intersect
	Touch
	Cross
	Within
	Contains
	Overlap
Spatial Analysis	Distance
	Buffer
	ConvexHull
	Intersection
	Union
	Difference
	DymmDiff

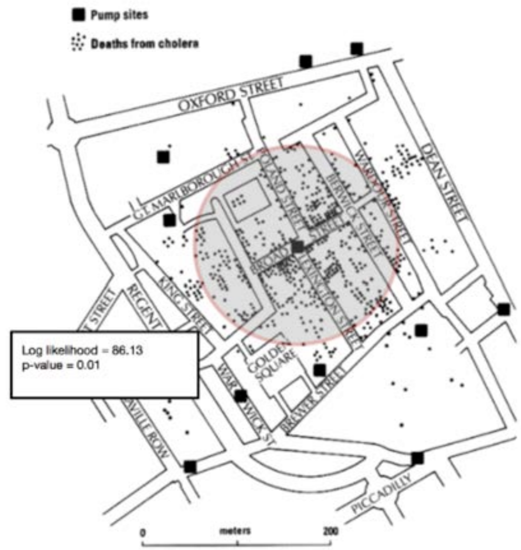
Details: Spatial Databases: Accomplishments and Research Needs, S. Shekhar et al., IEEE Trans. on Knowledge and Data Eng., 11(1), Jan.-Feb. 1999.

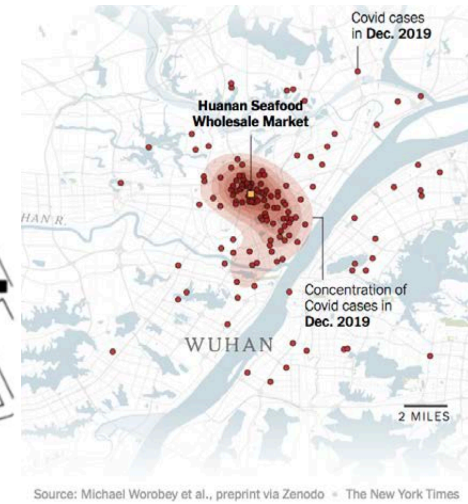
From Objects Detection to Pattern Mining

After object detection we look for Spatial patterns

- Hotspots, Spatial clusters
- Spatial outlier, discontinuities
- Co-locations, segregations
- Spatiotemporal predictions

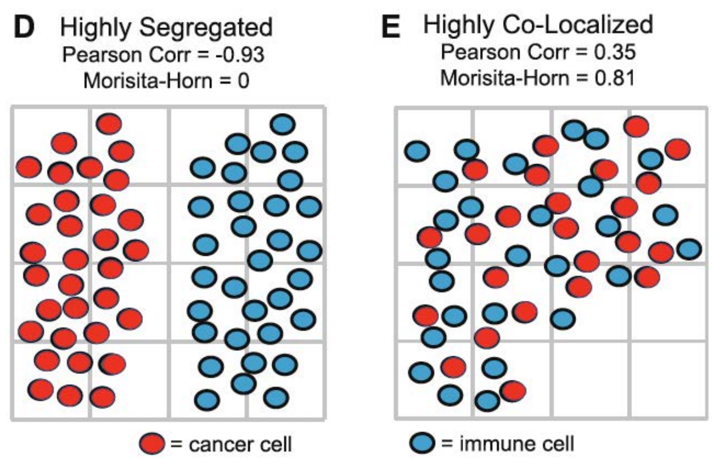
…





Spatial Data Mining is
The process of discovering

- interesting, useful, non-trivial patterns
- from large spatial datasets



Source: [An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer]. C. Maley et al., Breast Cancer Res 17, 131 (2015).

Spatial Pattern Mining Challenges

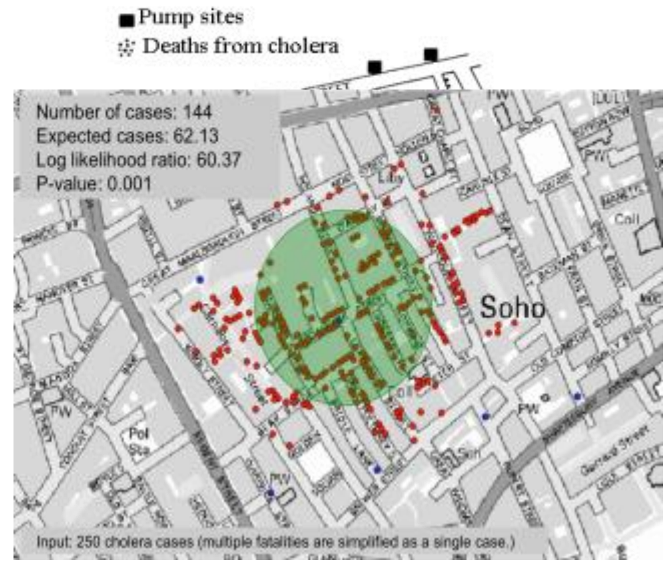
- Traditional pattern mining methods not robust in face of
  - Challenge 1: Noise
  - Challenge 2: Spatial continuity
  - Challenge 3: Auto-correlation, Heterogeneity , Edge-effect, …

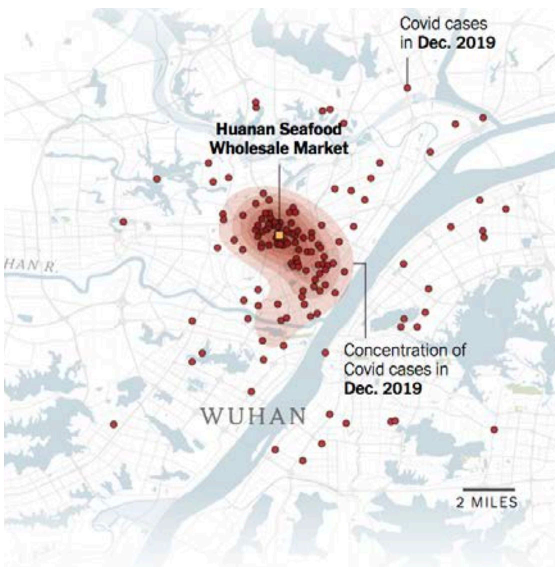
Details: Data Science for Earth: The Earth Day Report , E. Eftelioglu, S. Shekhar, J. Hudson, L. Joppa, C. Baru, V. Janeja, et al. ACM SIGKDD Explorations Newsletter, 22(1), May 2020

Dealing with Noise & Chance Patterns

- Statistics: Deal with Noise
  - Quantify uncertainty, confidence, …
  - Is it (statistically) significant?
  - Is it different from a chance event or rest of dataset?
    - e.g., SaTScan finds circular hot-spots

- Spatial Statistics, Spatial Data Mining
  - Auto-correlation, Heterogeneity, Edge-effect, …





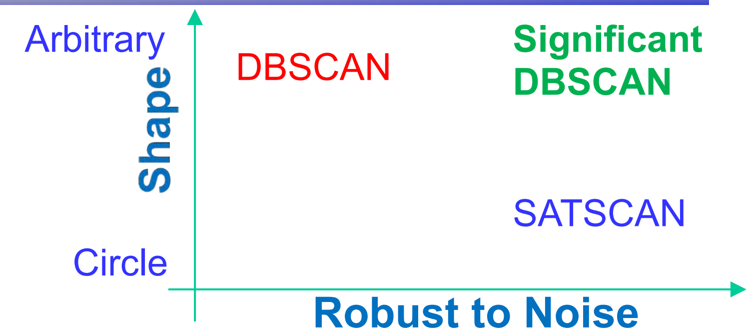
Source: Michael Worobey et al., preprint via Zenodo · The New York Times

SaTScan™

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

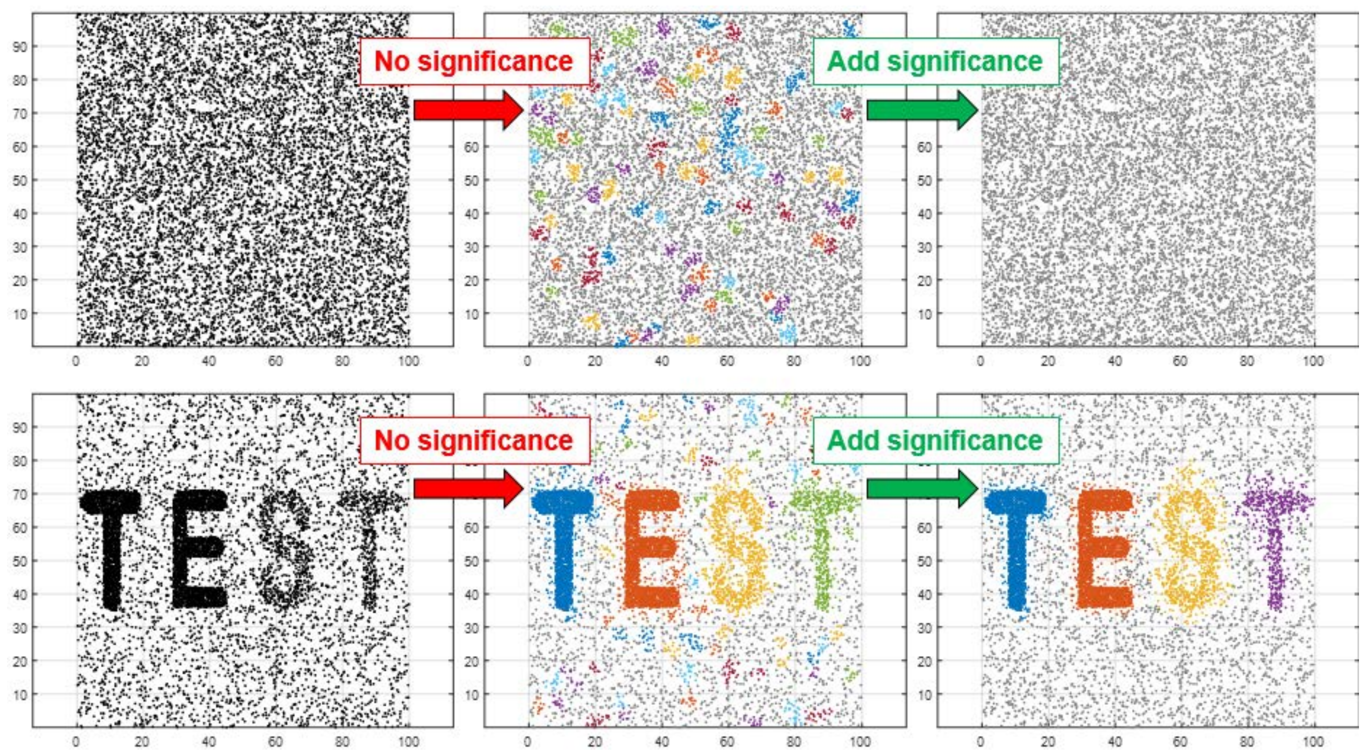
Robust Clustering (Hotspot Detection)

- **Problem definition**
  - **Inputs:** Collection of event locations, Test statistic; Significance level
  - **Output:** Significant clusters (hotspots)
  - **Constraints:** Avoid chance patterns despite non-trivial noise in data



Complete Spatial Random (noise)

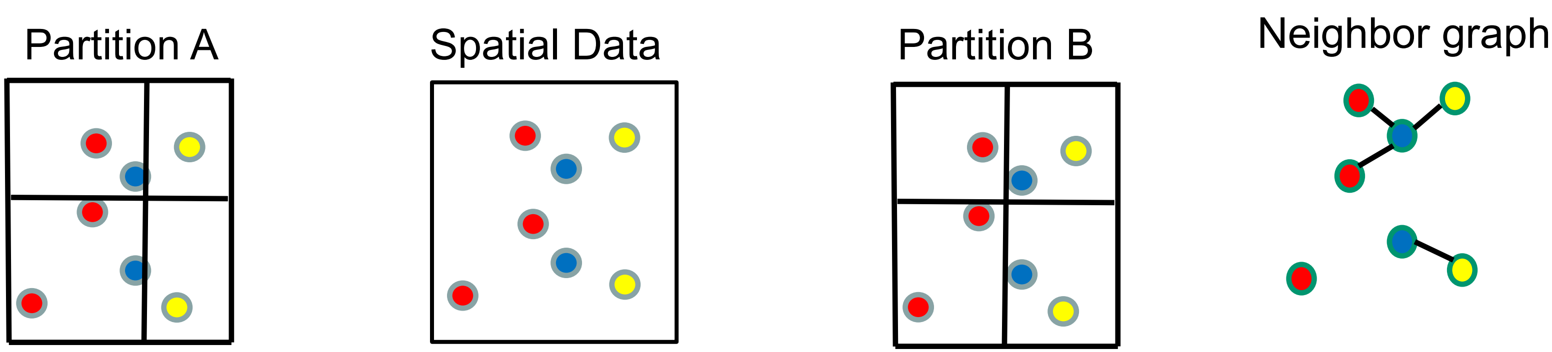
Hotspots with Noise



Details: Significant DBSCAN towards Statistically Robust Clustering, [ACM Trans. on Intelligent Systems and Tech](), 12(5):1-26, Oct. 2021. (A summary in 16th Intl. Symp. on Spatial and Temporal Databases, 2019. **(Best Paper Award)**

Challenge 2: Continuous Space

- Traditional relationship mining methods not robust
  - Result changes if spatial partitioning changes
  - Similar to Gerrymandering risk, Formally, Modifiable Areal Unit Problem (MAUP)
  - Neighbor Graph Based Measures are more robust



Partition A Based Pearson's Correlation	Pairs	Partition B Based Pearson's Correlation	Ripley's Cross-K	Participation Index
1	🔴 - 🔵	- 0.90	0.33	0.66
- 0.90	🟡 - 🔵	1	0.5	1

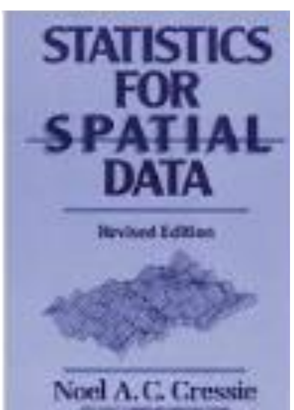
Details: Data Science for Earth: The Earth Day Report , E. Eftelioglu, S. Shekhar, J. Hudson, L. Joppa, C. Baru, V. Janeja, et al. ACM SIGKDD Explorations Newsletter, 22(1), May 2020

A Metric of Spatial Cross-Correlation

- Ripley's Cross K-Function Definition

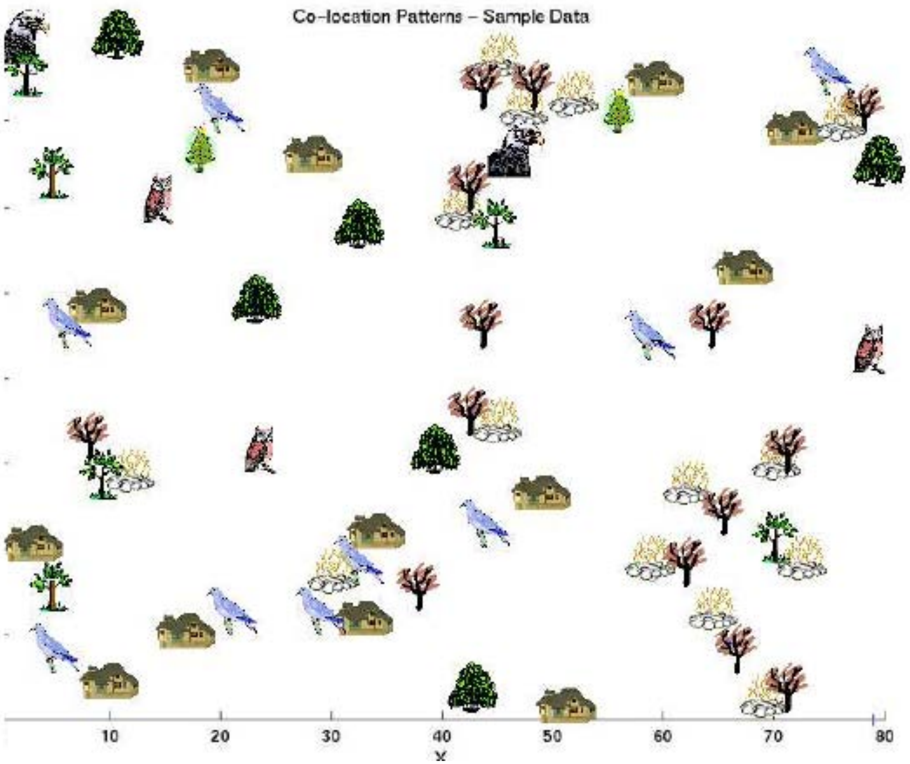
$$K_{ij}(h) = \lambda_j^{-1} E \text{ [number of type } \boldsymbol{j} \text{ event within distance } h \text{ of a randomly chosen type } \boldsymbol{i} \text{ event]}$$

  - Interpretation
    - Which pairs are frequently co-located
    - Statistical significance



Co-locations

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

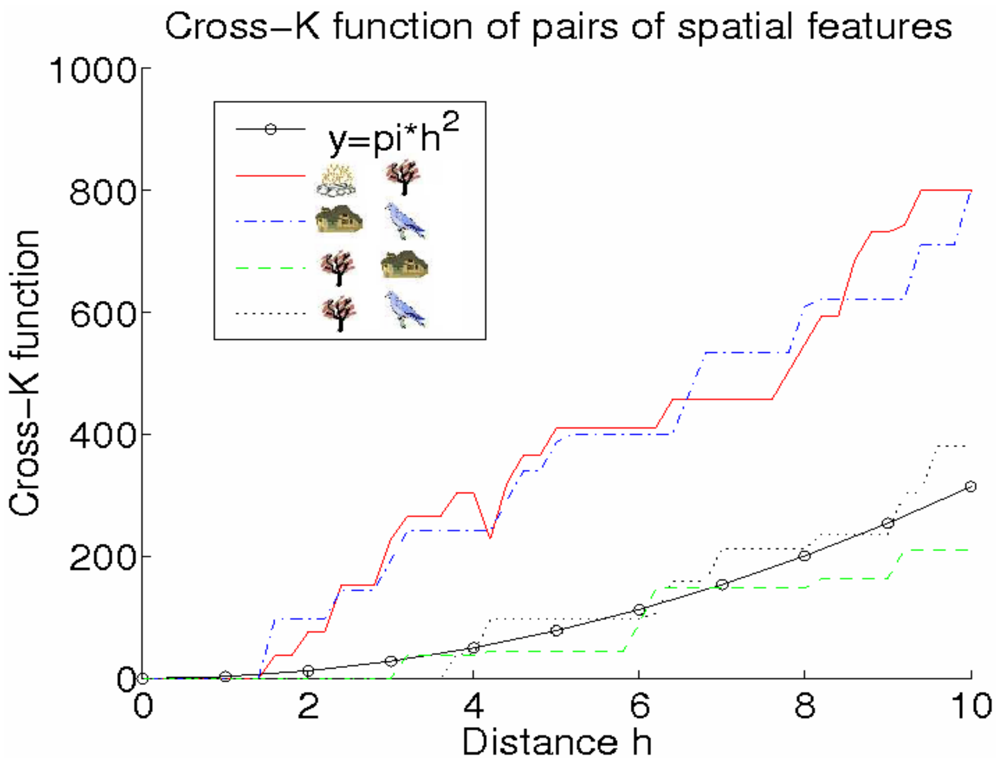


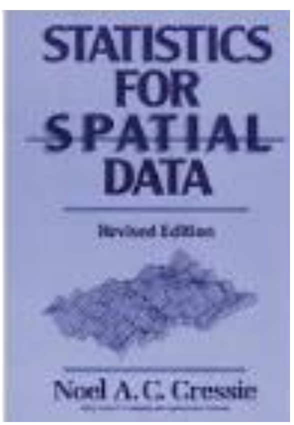
Source: Discovering Spatial Co-location Patterns: A General Approach, IEEE Transactions on Knowledge and Data Eng., 16(12), December 2004 (w/ H.Yan, H.Xiong).

Spatial Computing Research Group

Illustration of Cross-Correlation

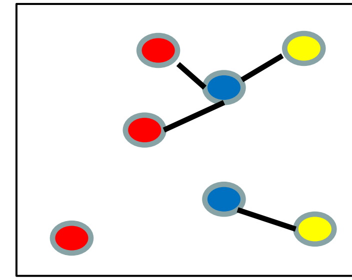
- Illustration of Cross K-function for Example Data





Spatial Computing Research Group

Spatial Colocation



Feature set: (, ,)

Feature Subsets:

Participation ratio (pr):

pr(,) = fraction of instances neighboring feature { } = 2/3

pr(,) = ½

Participation index ( ) = **pi**( )

= min{ **pr**(,), **pr**(,) }

= min (2/3, ½) = ½

Participation Index Properties:

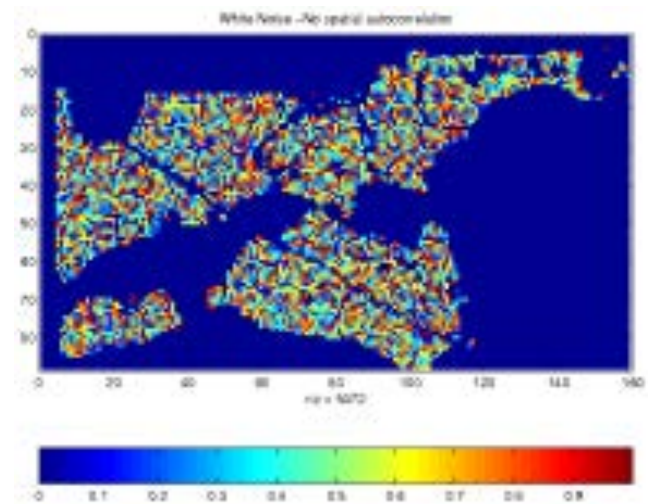
(1) Computational: Non-monotonically decreasing like support measure

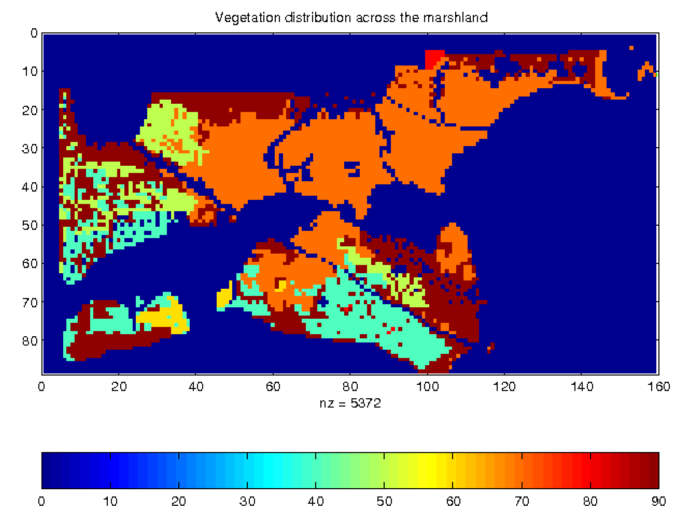
(2) Statistical: Upper bound on Ripley's Cross-K function

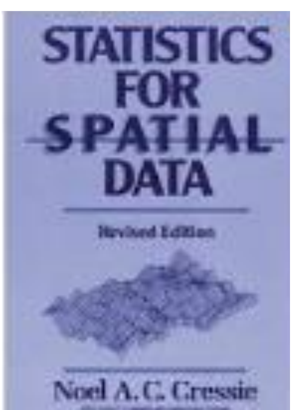
Details: Discovering Spatial Co-location Patterns: A General Approach, IEEE Transactions on Knowledge and Data Eng., 16(12), December 2004 (w/ H. Yan, H. Xiong).

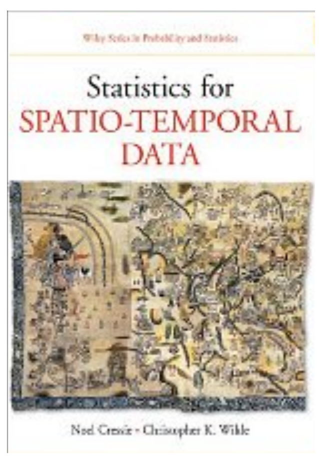
Challenge 3: Spatial Auto-correlation

- Traditional Statistics, ML, Data Mining
- Ubiquitous i. i. d. assumption
  - Data samples independent of each other
  - From identical distribution
- Problem
  - Ignores auto-correlation, heterogeneity
  - Salt n Pepper noise





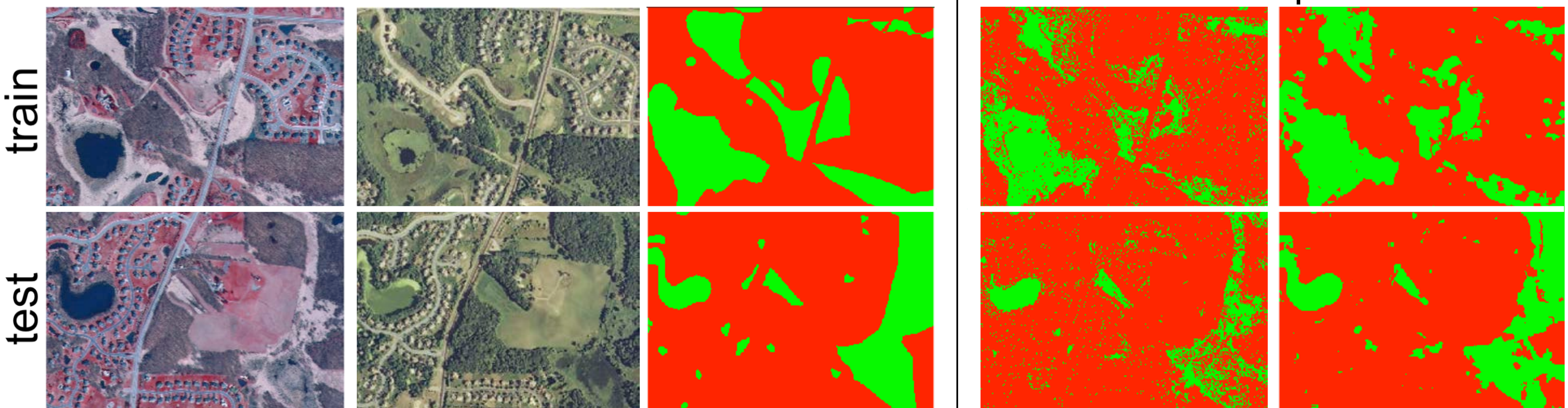




Ex. Salt n Pepper Noise

wetland dry land

Input: Output:



(a) aerial photo (b) aerial photo (c) true classes (d) DT prediction (Salt n Pepper Noise) (e) SDT prediction

Training samples: upper half
Test samples: lower half
Spatial neighborhood: maximum 11 pixels by11 pixels

DT: decision tree
SDT: spatial decision tree

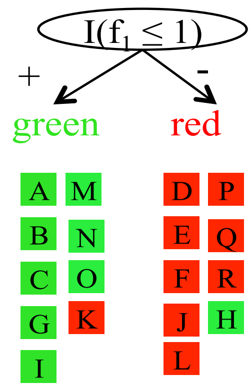
Details: Focal-Test-Based Spatial Decision Tree Learning. IEEE Trans. Knowl. Data Eng. 27(6): 1547-1559, 2015 (A summary in Proc. IEEE Intl. Conf. on Data Mining, 2013).(w/ Z. Jiang et al.)

Spatial Decision Tree

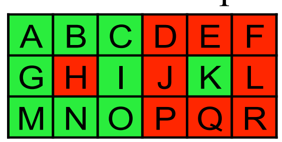
Traditional decision tree

Inputs: table of records

ID	f_1	f_2	Γ_1	class
A	1	1	1	green
B	1	1	0.3	green
C	1	3	0.3	green
G	1	1	0.3	green
I	1	3	0	green
K	1	2	-1	red
M	1	1	1	green
N	1	1	0.3	green
O	1	3	0.3	green
D	3	2	0.3	red
E	3	2	0.3	red
F	3	2	1	red
H	3	1	-1	green
J	3	2	0	red
L	3	2	0.3	red
P	3	2	0.3	red
Q	3	2	0.3	red
R	3	2	1	red



Predicted map



feature test	information gain
$f_1 \le 1$	0.50
$f_2 \le 1$	0.46
$f_2 \le 2$	0.19

Spatial decision tree

Inputs:

- feature maps, class map
- Rook neighborhood

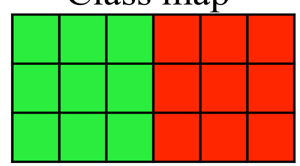
Feature f_1

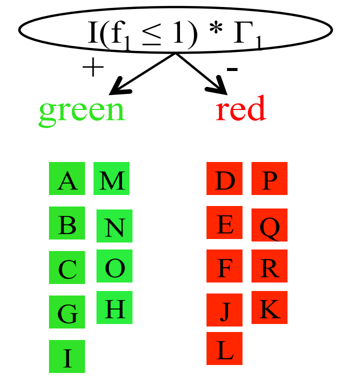
1	1	1	3	3	3
1	3	1	3	1	3
1	1	1	3	3	3

Feature f_2

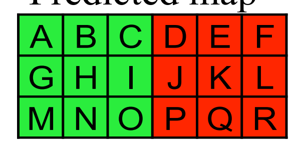
1	1	3	2	2	2
1	1	3	2	2	2
1	1	3	2	2	2

Class map





Predicted map



Modeling Spatial Auto-correlation

- Traditional Models, e.g., Regression (with Logit or Probit),
  - Linear Regression (LR), Bayes Classifier, …
- Semi-Spatial: LR with auto-corr. regularizer $\varepsilon = \|y - \beta X\|^2 + \|\beta X - \beta X_{neighbor}\|^2$
- Spatial
  - Spatial autoregressive model (SAR)
  - Markov random field (MRF) based Bayesian Classifier

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

Spatial Computing Research Group

Spatial Auto-Regression & Parameter Estimation

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

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

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

- **Maximum Likelihood Estimation**

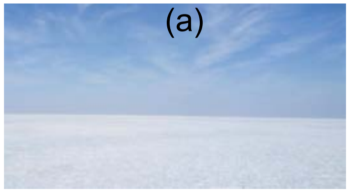
$$\ln(L) = \boxed{\ln\left|\mathbf{I} - \rho\mathbf{W}\right|} - \frac{n\ln(2\pi)}{2} - \frac{n\ln(\sigma^2)}{2} - SSE$$

- 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.

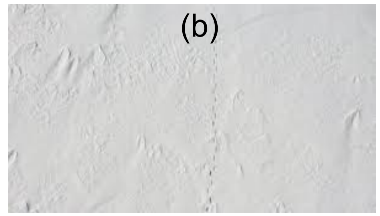
Details: A parallel formulation of the spatial auto-regression model for mining large geo-spatial datasets, SIAM Intl. Workshop on High Perf. and Distr. Data Mining, 2004. (with B. Kazar)

Spatial Heterogeneity

- Knowledge of location can improve land-cover and object recognition
  - Q? Which pictures show snow?



Salt Marsh
(Runn of Kutch, Gujarat, India)



Snow

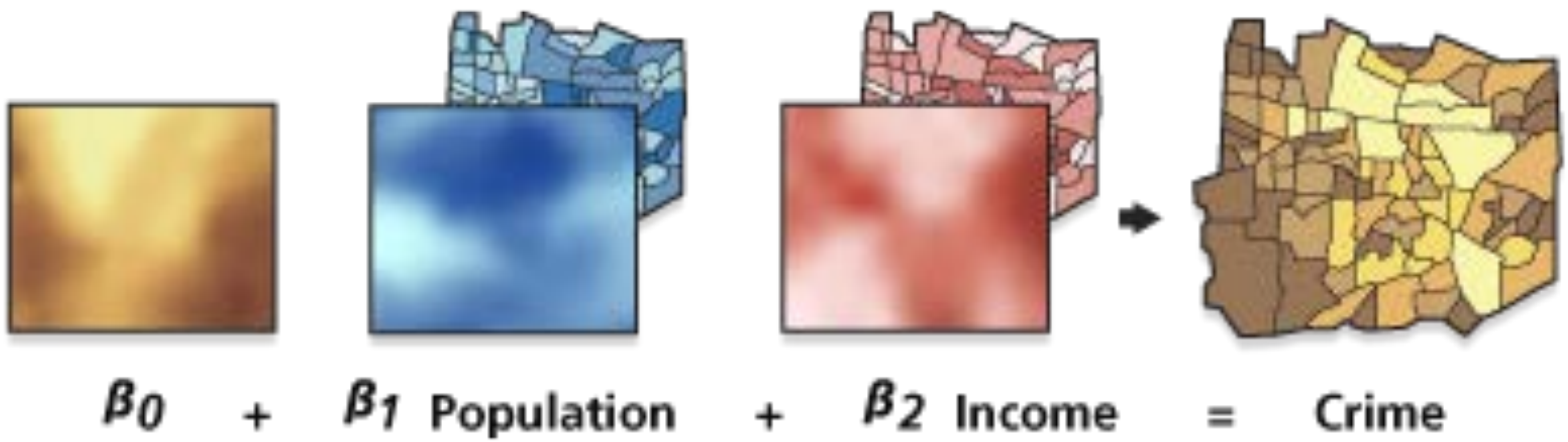
Snow

- Coarse Satellite Imagery (e.g., 30m pixels)
  - More effective for large mono-crop farms the small mixed-crop plots

- However, Convolutional Neural Networks does not model geographic heterogeneity.

Modeling Spatial Heterogeneity: GWR

- Geographically Weighted Regression (GWR)
  - Goal: Model spatially varying relationships
  - Example: $y = X\beta' + \varepsilon'$
    Where β' and ε' are location dependent



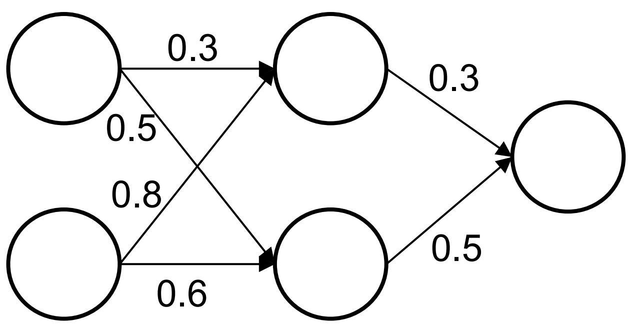
Source: resources.arcgis.com

Spatial Computing Research Group

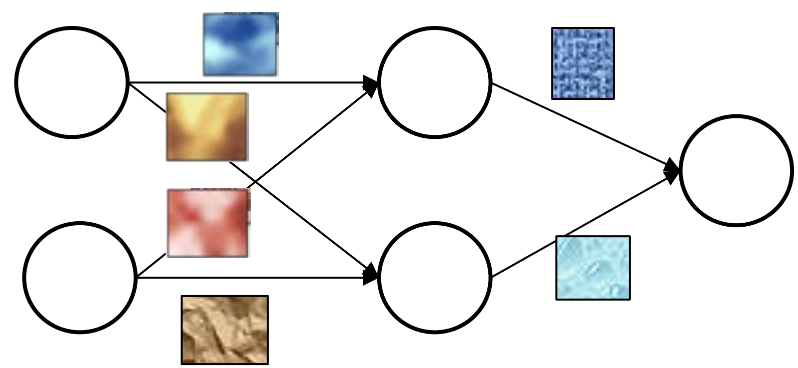
Spatial Variability Aware Neural Networks (SVANN)

- Each NN parameter is a map i.e., a function of location
  - Similar to Geographically Weighted Regression

A Neural Network (NN)



SVANN



- Evaluation:
  - Urban Garden Detection across Hennepin County, MN and Fulton County, GA.
  - SVANN outperformed OSFA by 14.34% on F1-scores.

Details: Towards Spatial Variability Aware Deep Neural Networks (SVANN), [ACM Transactions on Intelligent Systems and Technology](), 12(6):1-21, Dec. 2021. (A Summary in ACM SIGKDD Workshop on Deep Learning for Spatiotemporal Data, Applications, and Systems, 2020. (Best Paper Award)

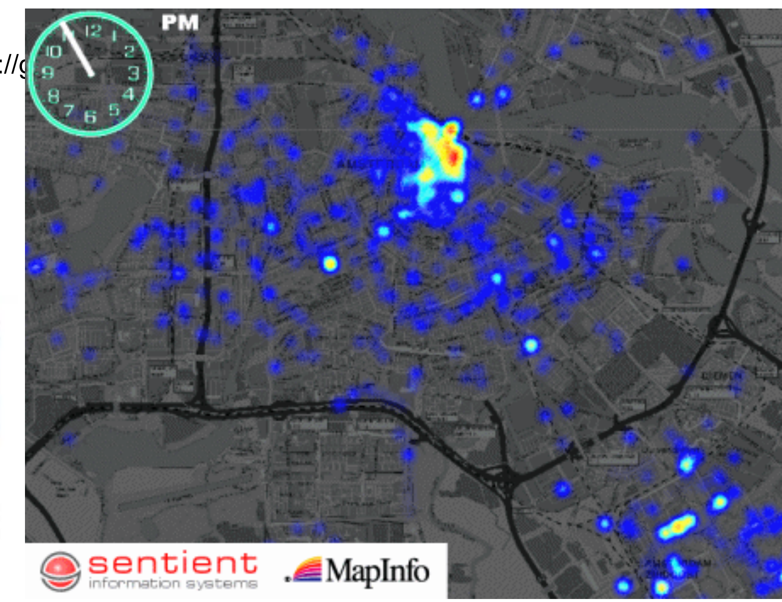
What has changed?

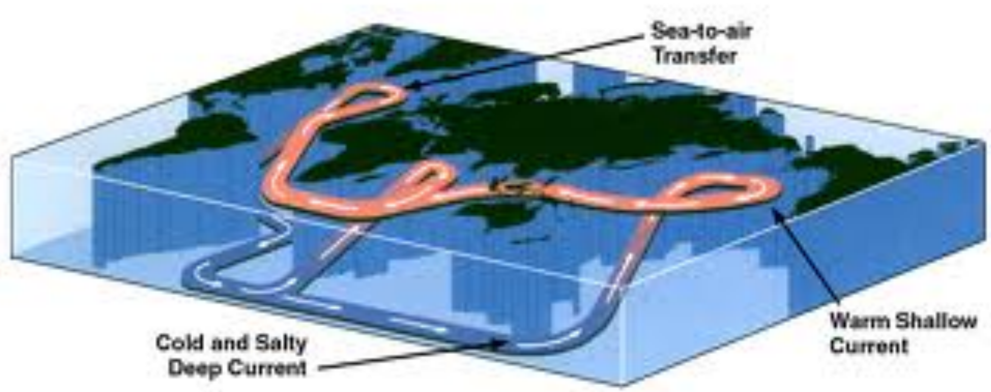
	Last Century	This Century
Spatial Data Revolution	Smaller Data	Spatial Big Data
Spatial Data Access and Platforms	Smaller platforms	Big Compute, e.g., Cloud (e.g., AWS Earth, ESRI GIS Tools for Hadoop)
Spatial Data Processing	Fairly manual, labor-intensive (Geo-Intelligence)	More automation Geo-Augmented-Intelligence (Geo-AI)
Spatial Data Science	“One-size fit all” AI applied to spatial data	Virtuous cycle between Geo and AI
Spatial Data Visualization	Maps, albums	Spatio-temporal, 3D

Towards 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 30+frame video
  - https://earthengine.google.com/timelapse/
  - Dubai coastal expansion
  - Chicago O'Hare airport
  - Doha, Qatar
  - Marina Center, Singapore (Wikipedia entry)
  - Salt Lake, Bidhannagar, Kolkata, WB, India
  - UMN, Minneapolis;airport, MN, USA

http://g … html





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

Data that are geographically referenced or contain some type of location markers are both common and of high value (e.g., data subject to state-specific policies, laws and regulations; demographic data from the census; location traces of smartphones and vehicles; remotely sensed imagery from satellites, aircraft and small unmanned aerial vehicles; volunteered geographic information; geographically referenced social media postings). A 2011 McKinsey Global Institute report estimates a value of "about $600 billion annually by 2020" from leveraging personal location data[2] to reduce fuel waste, improve health outcomes, and better match products to consumer needs. Spatial data are critical for societal priorities such as national security, public health & safety, food, energy, water, smart cities, transportation, climate, weather, and the environment. For example, remotely-sensed satellite imagery is used to monitor not only weather and climate but also global crops[3] for early warnings and planning to avoid food shortages.

University Consortium for GEOGRAPHIC INFORMATION SCIENCE

Summer 2018

One Size Data Science Does not Fit All Data!

However, spatial data presents unique data science challenges. Recent court cases that address gerrymandering, the manipulation of geographic boundaries to favor a political party, offer a high-profile example. Instances of such exploitation of the modifiable areal unit problem (or dilemma) is not limited to elections since the MAUP affects almost all traditional data science methods in which results (e.g., correlations) change dramatically by varying geographic boundaries of spatial partitions. The fundamental geographic qualities of spatial autocorrelation, which assumes properties of geographically proximate places to be similar, and geographic heterogeneity, where no two places on Earth are exactly alike, violate assumptions of sample independence and randomness that underlie many conventional statistical methods. Other spatial challenges include how to choose between a plurality of projections and coordinate systems and how to deal with the imprecision, inaccuracy, and uncertainty of location

A UCGIS Call to Action:

Bringing the Geospatial Perspective to Data Science Degrees and Curricula

University Consortium for
GEOGRAPHIC INFORMATION SCIENCE

Summer 2018

Spatial Data Science Tools

measurements. To deal with such challenges, practitioners in many fields including agriculture, weather forecast, mining, and environmental science incorporate *geospatial data science*[4] methods such as spatially-explicit models, spatial statistics[5], geo-statistics, geographic data mining[6], spatial databases[7], etc.

[4] Y. Xie et al., Transdisciplinary Foundations of Geospatial Data Science, *ISPRS Intl. Jr. of Geo-Informatics*, 6(12):395-418, 2017. DOI: 10.3390/ijgi6120395.

[5] N. Cressie, *Statistics for Spatial Data*, Wiley, 1993 (1st ed.), 2015 (Revised ed.).

[6] H. Miller and J. Han, *Geographic Data Mining and Knowledge Discovery*, CRC Press, 2009 (2nd Ed.).

[7] S. Shekhar and S. Chawla, *Spatial Databases: A Tour*, Prentice Hall, 2003.

A UCGIS Call to Action:

Bringing the Geospatial Perspective to Data Science Degrees and Curricula

University Consortium for GEOGRAPHIC INFORMATION SCIENCE

Summer 2018

Summary

The World Economy
Runs on GPS.

- Spatial Data has already transformed our society
  - It is only a beginning!
  - It promises astonishing opportunities in coming decade

- AI has promise but faces major challenges
  - Rich Data Types, e.g., lineStrings, polygons, …
  - High cost of errors, Spatial Heterogeneity, …
- **Ask**
  - Sponsors: Nurture approaches to overcome challenges (Geo-AI
  - Academics: Include Spatial topics in courses and curricula

University Consortium for
GEOGRAPHIC INFORMATION SCIENCE
Summer 2018

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

References :Surveys, Overviews

- Spatial Computing, MIT Press (Essential Knowledge Series), 2020. (ISBN: 9780262538046).
- Spatial Computing, Communications of the ACM, 59(1):72-81, January, 2016.
- Transdisciplinary Foundations of Geospatial Data Science, 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, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3):193-214, May/June 2011.
- 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.
- 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.