Spatial Artificial Intelligence: Introduction

Yao-Yi Chiang
Computer Science and Engineering
University of Minnesota
yaoyi@umn.edu

What is Spatial Artificial Intelligence?

OCT 2021

https://datareportal.com/global-digital-overview

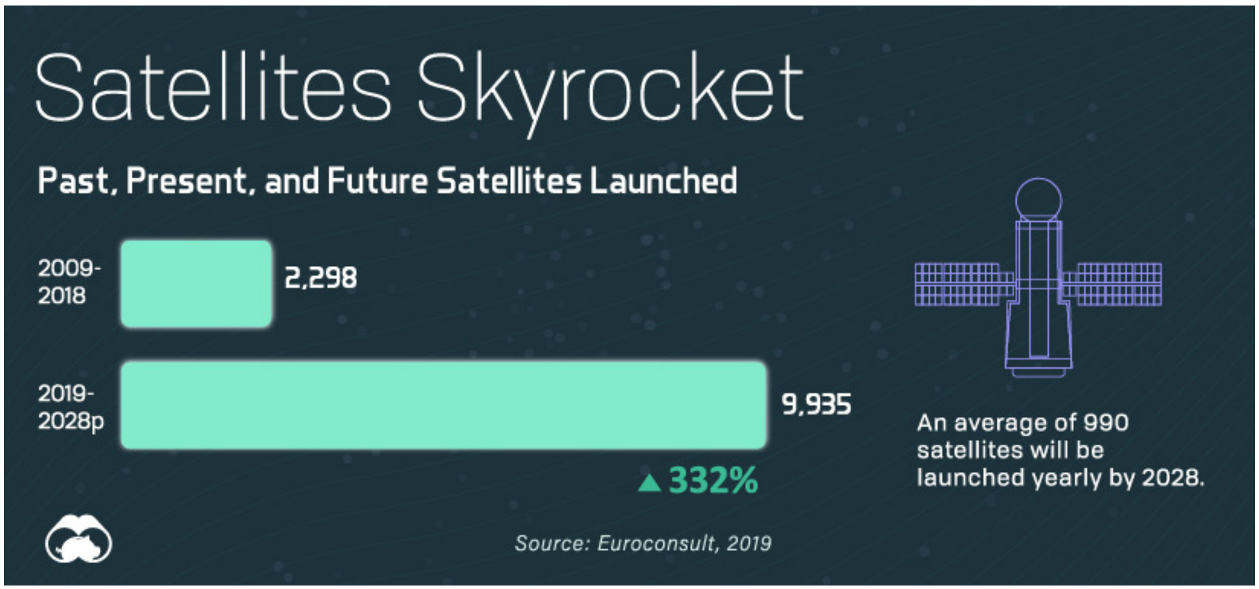
DIGITAL AROUND THE WORLD

ESSENTIAL HEADLINES FOR MOBILE, INTERNET, AND SOCIAL MEDIA USE

INTERNET USER NUMBERS NO LONGER INCLUDE DATA SOURCED FROM SOCIAL MEDIA PLATFORMS, SO VALUES ARE NOT COMPARABLE WITH PREVIOUS REPORTS

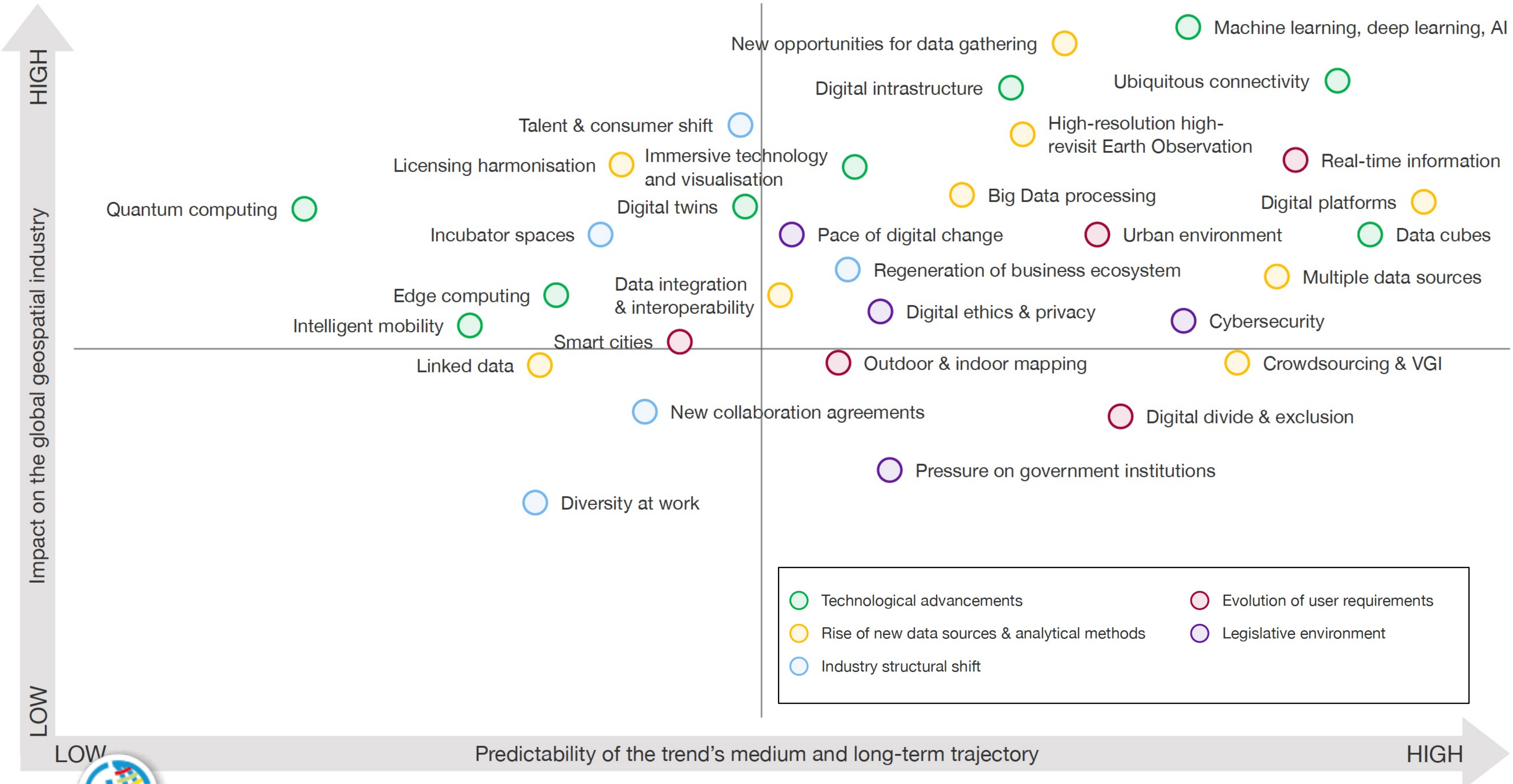
TOTAL POPULATION	UNIQUE MOBILE PHONE USERS	INTERNET USERS*	ACTIVE SOCIAL MEDIA USERS*
7.89 BILLION	5.29 BILLION	4.88 BILLION	4.55 BILLION
URBANISATION: 56.6%	vs. POPULATION: 67.1%	vs. POPULATION: 61.8%	vs. POPULATION: 57.6%

SOURCES: THE U.N.; LOCAL GOVERNMENT BODIES; GSMA INTELLIGENCE; ITU; GWI; EUROSTAT; CNNIC; APJII; SOCIAL MEDIA PLATFORMS' SELF-SERVICE ADVERTISING TOOLS; COMPANY EARNINGS REPORTS; MEDIASCOPE. ***ADVISORIES:** INTERNET USER NUMBERS NO LONGER INCLUDE DATA SOURCED FROM SOCIAL MEDIA PLATFORMS, SO VALUES ARE NOT COMPARABLE TO DATA PUBLISHED IN PREVIOUS REPORTS. SOCIAL MEDIA USER NUMBERS MAY NOT REPRESENT UNIQUE INDIVIDUALS. ◈ **COMPARABILITY ADVISORY:** SOURCE AND BASE CHANGES.



https://www.visualcapitalist.com/visualizing-all-of-earths-satellites/

https://twitter.com/TheEconomist/status/860135249552003073/photo/1



Machine learning, deep learning, AI
New opportunities for data gathering
Digital intrastructure
Ubiquitous connectivity
Talent & consumer shift
High-resolution high-revisit Earth Observation
Licensing harmonisation
Immersive technology and visualisation
Real-time information
Quantum computing
Digital twins
Big Data processing
Digital platforms
Incubator spaces
Pace of digital change
Urban environment
Data cubes
Regeneration of business ecosystem
Multiple data sources
Edge computing
Data integration & interoperability
Digital ethics & privacy
Cybersecurity
Intelligent mobility
Smart cities
Linked data
Outdoor & indoor mapping
Crowdsourcing & VGI
New collaboration agreements
Digital divide & exclusion
Pressure on government institutions
Diversity at work
Technological advancements
Rise of new data sources & analytical methods
Industry structural shift
Evolution of user requirements
Legislative environment
HIGH
Impact on the global geospatial industry
LOW
LOW
Predictability of the trend's medium and long-term trajectory
HIGH

What is Artificial Intelligence?

Artificial Intelligence

Broadly speaking, any technologies having human-like capabilities to perform certain tasks

Data Mining

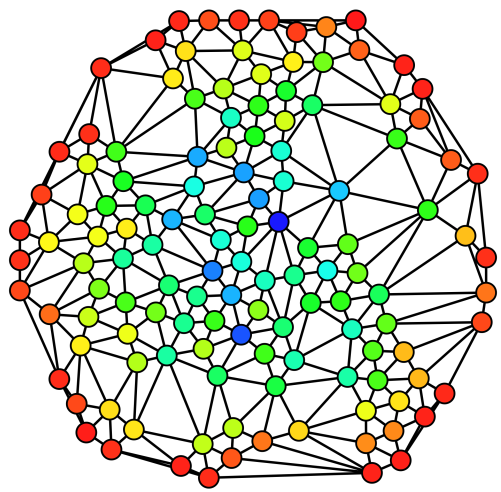
Discovering previously unknown knowledge from BIG data

Machine Learning

(Machine) Making inference using available data without explicit rules

AI Tasks - Descriptive Analysis

Detect previously unknown patterns in data



e.g., split social networks into groups

https://en.wikipedia.org/wiki/Social_network_analysis#/media/File:Graph_betweenness.svg

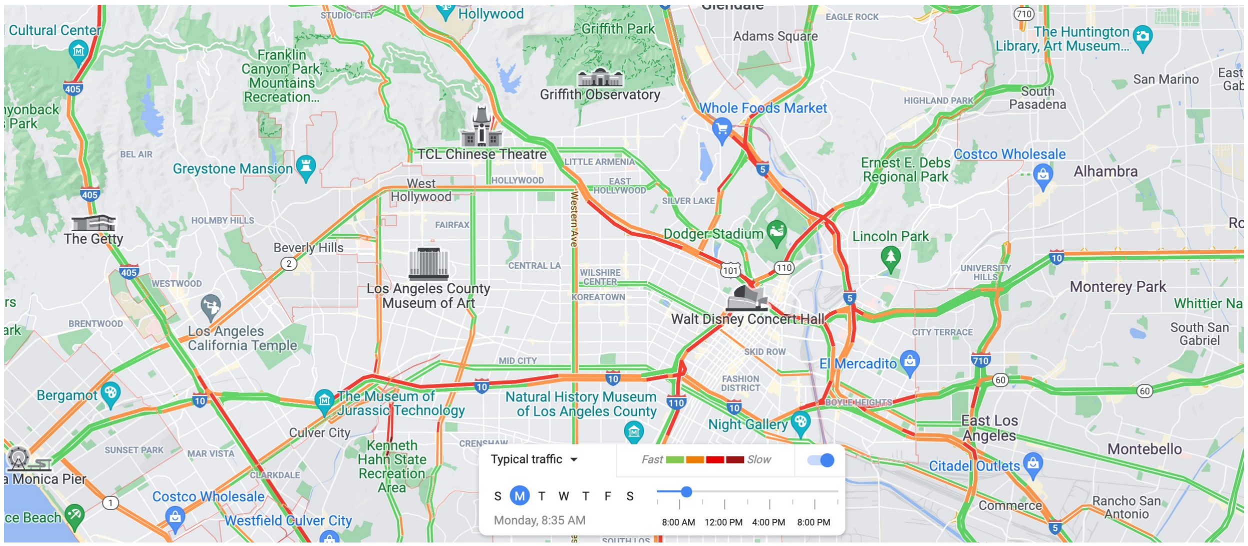
AI Tasks - Predictive Analysis

Predict unobserved data values using computer models

e.g., by learning some model parameters from some data to predict future stock prices

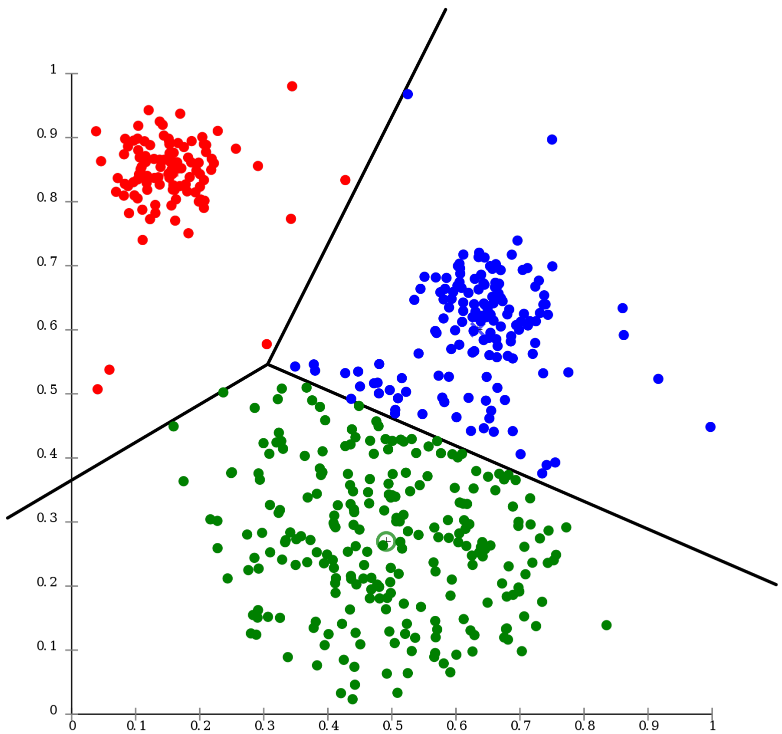
AI Tasks - Prescriptive Analysis

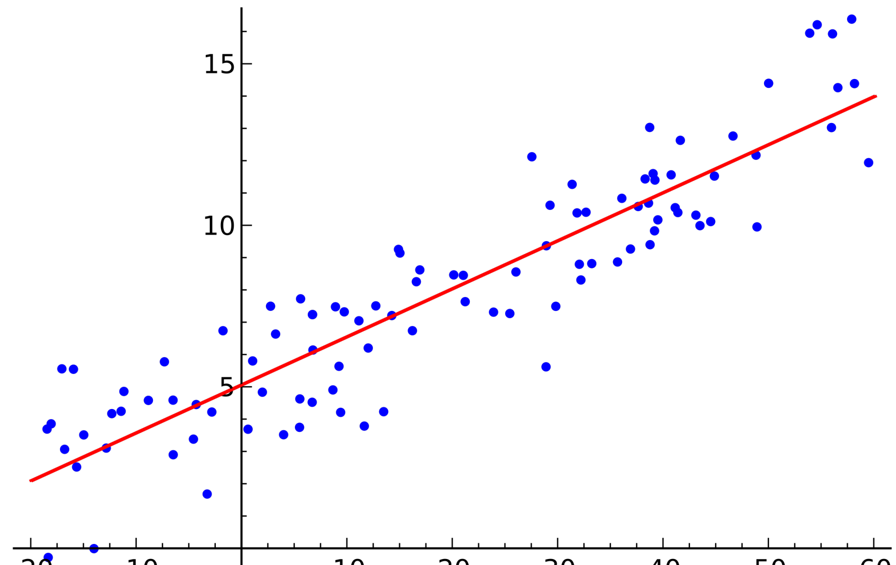
Examinate “What-If” scenarios and suggest actions



e.g., traffic on I-10 will be slow in the next hour, take I-210 will save 15 minutes

Patterns vs. Models

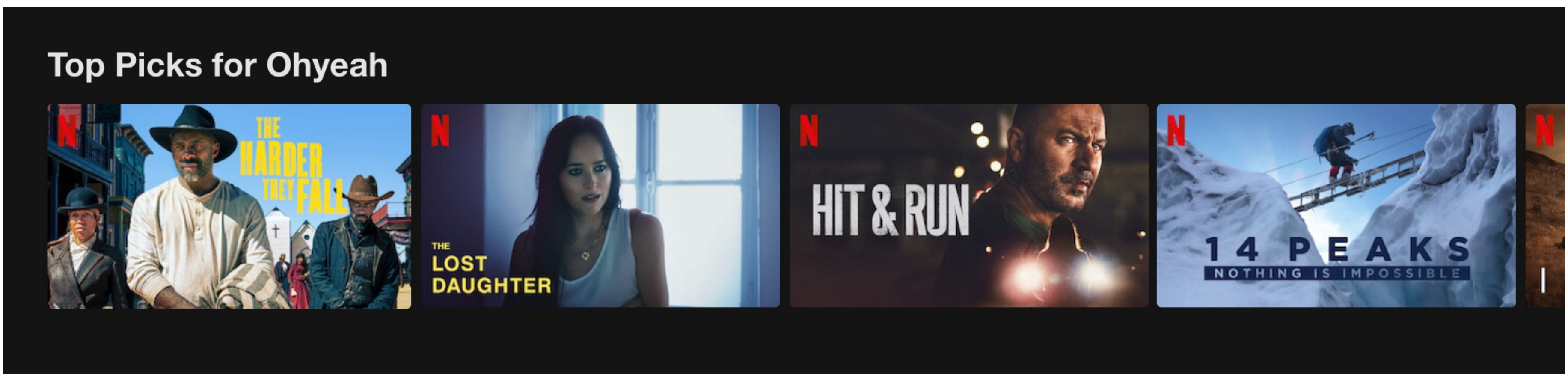




https://en.wikipedia.org/wiki/Cluster_analysis#/media/File:KMeans-Gaussian-data.svg
https://en.wikipedia.org/wiki/Machine_learning#/media/File:Linear_regression.svg

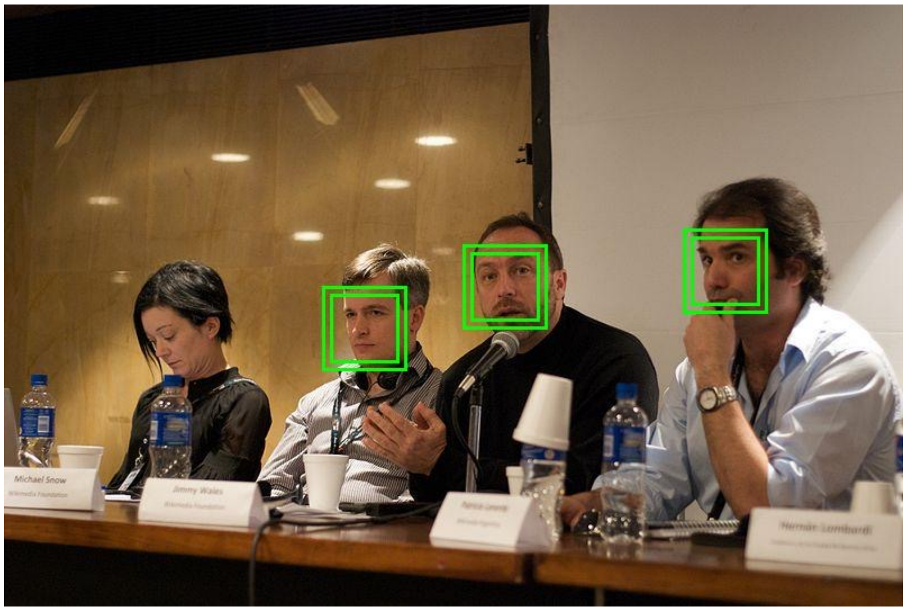
Data Mining Examples

- If you buy beer, you highly likely will buy diapers (association rules)
- People who have a similar profile as you like to watch these shows (recommendation systems)

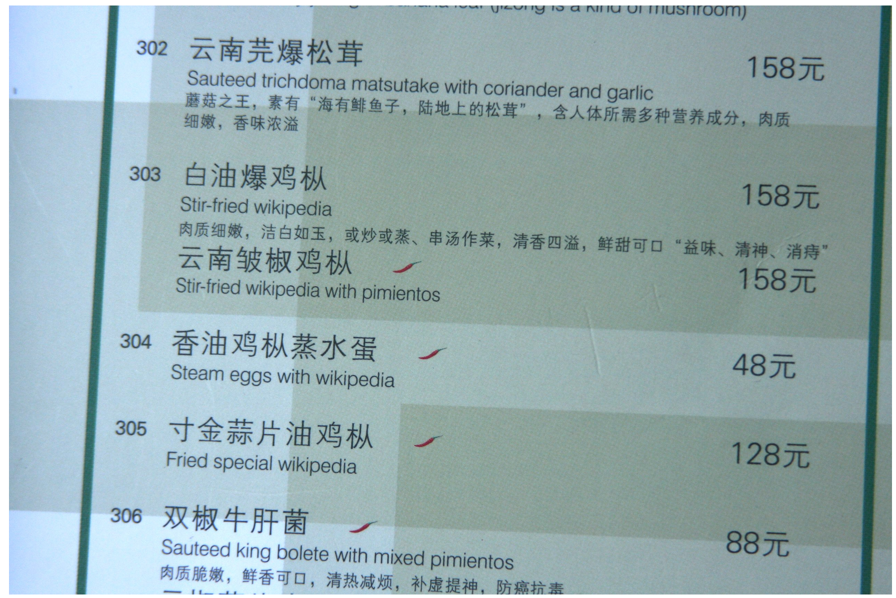


Machine Learning Examples

Face Detection (Computer Vision)



Machine Translation (Natural Language Processing)



https://en.wikipedia.org/wiki/Facial_recognition_system#/media/File:Face_detection.jpg
https://en.wikipedia.org/wiki/Machine_translation#/media/File:Stir_Fried_Wikipedia.jpg

What is Spatial AI?

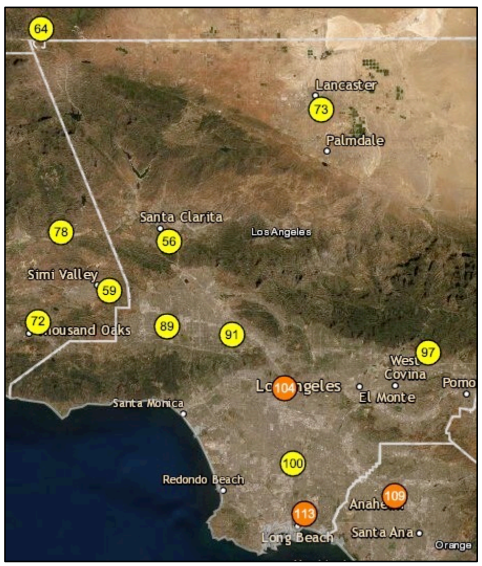
Spatial Artificial Intelligence

AI technologies that handle spatial data, typically associated with real-world applications

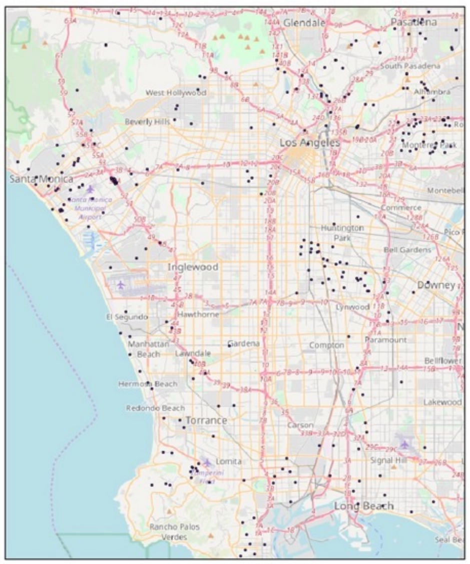
- Data Mining
- Machine Learning
- Geo-AI
- Spatial Statistics
- Geographic Information Science
- …

Spatial AI Example

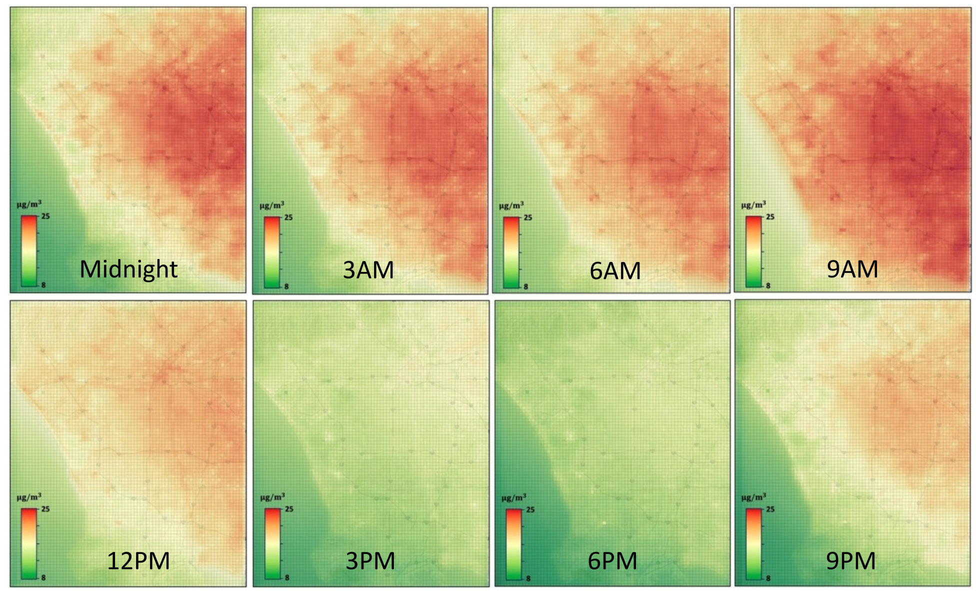
Air Quality Prediction



South Coast AQMD Monitors
https://gispub.epa.gov/airnow/



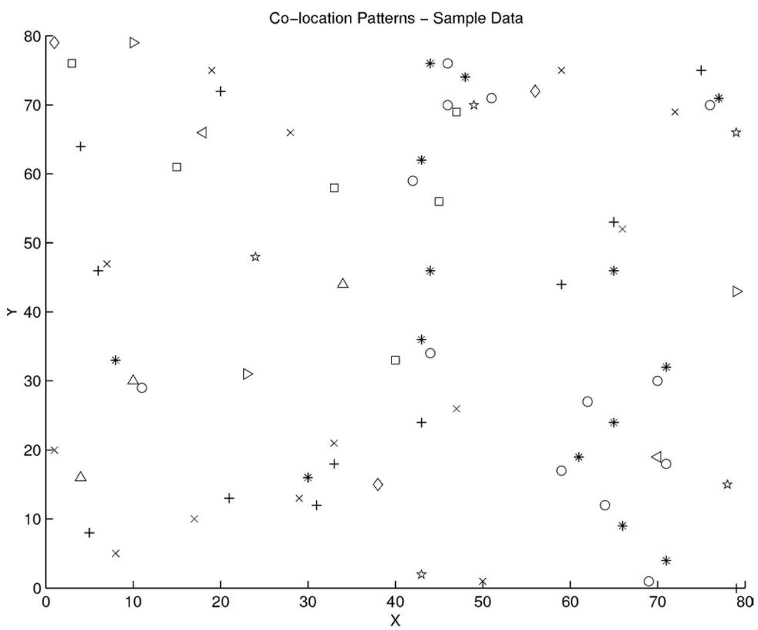
PurpleAir Sensors
https://www2.purpleair.com/
(50km×40km geographic area)



Lin, Y., Chiang, Y. Y., Franklin, M., Eckel, S. P., & Ambite, J. L. (2020, November). Building Autocorrelation-Aware Representations for Fine-Scale Spatiotemporal Prediction. In ICDM) (pp. 352-361)

Spatial AI Example

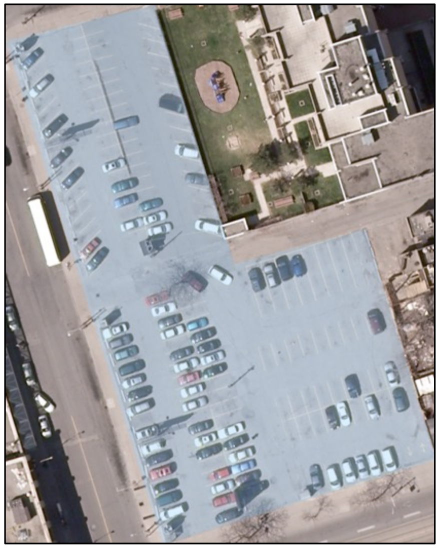
Spatial colocation mining



Huang, Y., Shekhar, S., & Xiong, H. (2004). Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and data engineering*, *16*(12), 1472-1485.

Spatial AI Example

Object detection from overhead imagery

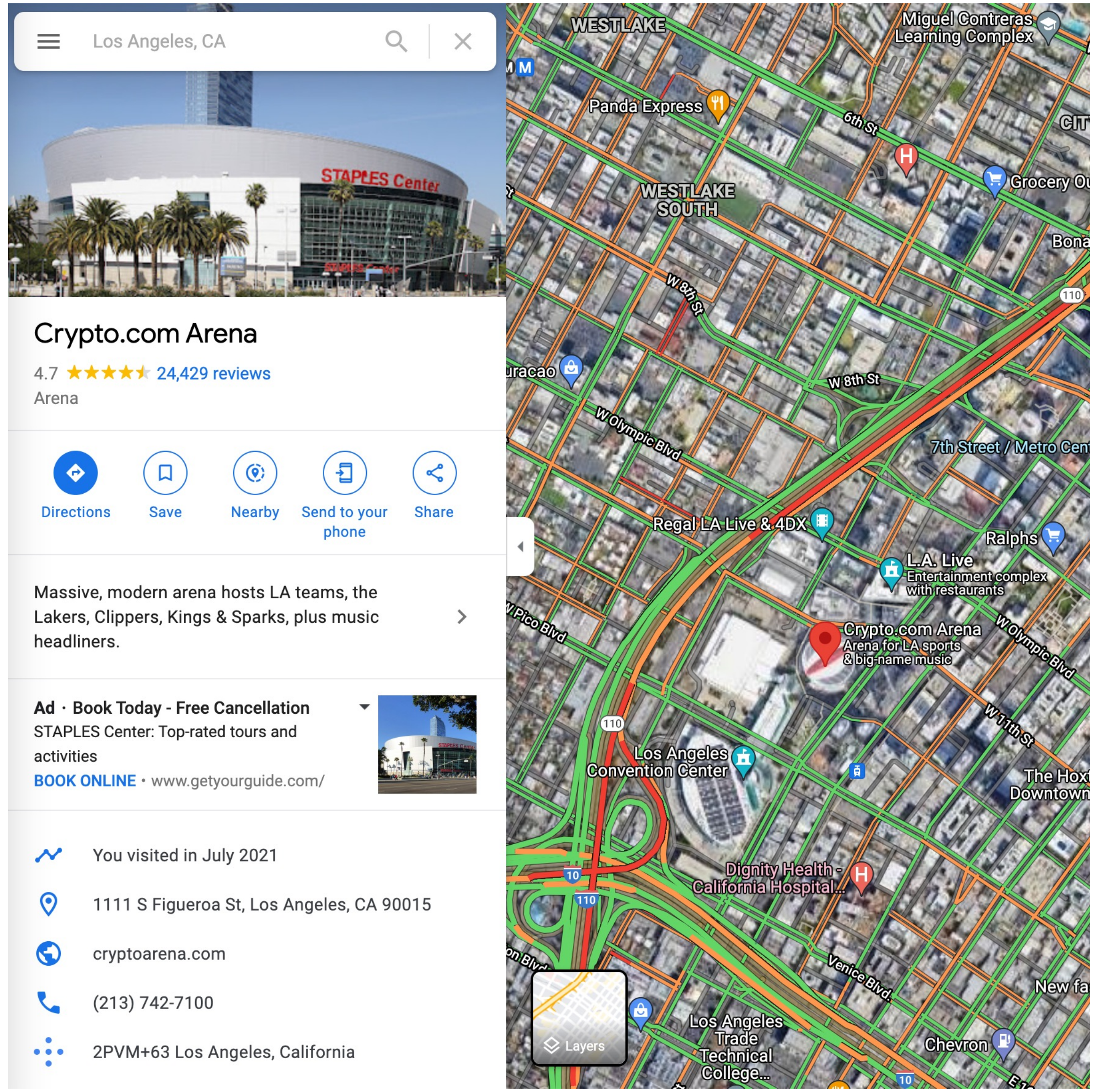


Duan, W., Chiang, Y.-Y., Leyk, S., Uhl, J. H., and Knoblock, C. A. (December 2021). Guided Generative Models using Weak Supervision for Detecting Object Spatial Arrangement in Overhead Images. IEEE Big Data (accepted), online

What are Spatial Data?

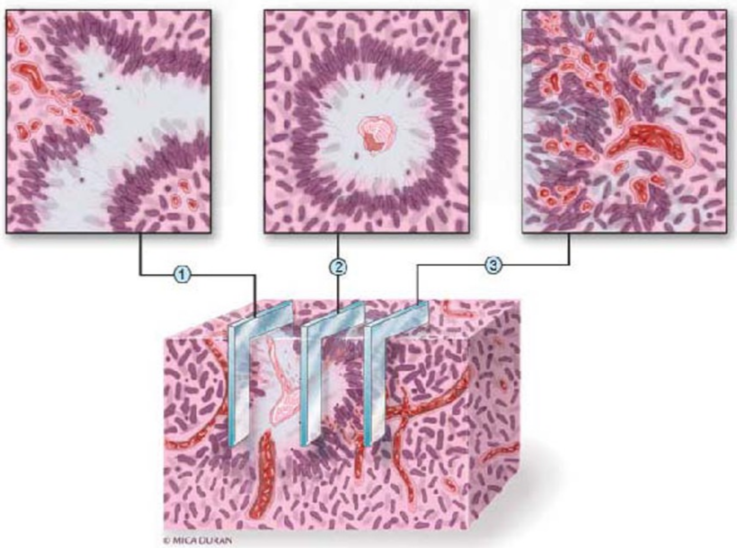
Spatial Data

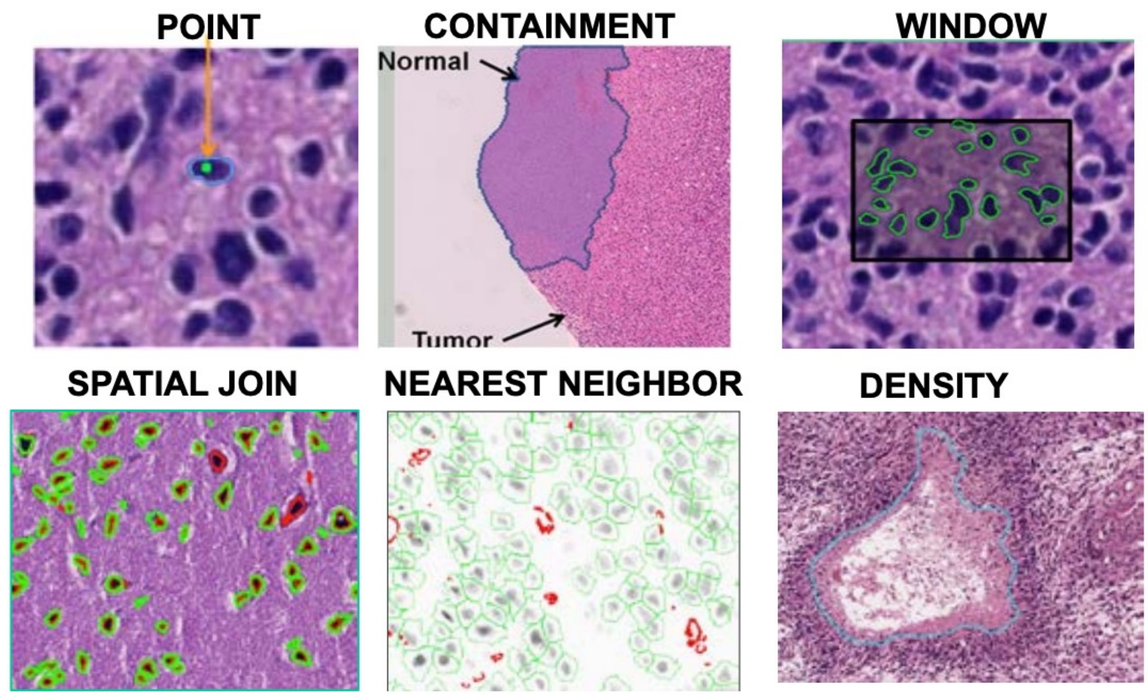
- Data that can spatially referenced, e.g.,
  - Time series from fixed-site sensors (e.g., traffic, air quality)
  - Remotely sensed data (e.g., satellite imagery)
  - Geotagged photos and tweets
  - Documents mentioning location entities



Spatial Data Do Not Have to be Geo Data

- Digital Pathology Example





Rong, Y., Durden, D. L., Van Meir, E. G., & Brat, D. J. (2006). 'Pseudopalisading'necrosis in glioblastoma: a familiar morphologic feature that links vascular pathology, hypoxia, and angiogenesis. *Journal of Neuropathology & Experimental Neurology*, *65*(6), 529-539.

Fox, Geoffrey & Qiu, Judy & Crandall, David & von Laszewski, Gregor & Jha, Shantenu & Wang, Fusheng & Marathe, Madhav & Paden, J. & Cheatham, Tom & Beckstein, Oliver. (2016). Datanet: CIF21 DIBBs: Middleware and High Performance Analytics Libraries for Scalable Data Science NSF14-43054 Progress Report.

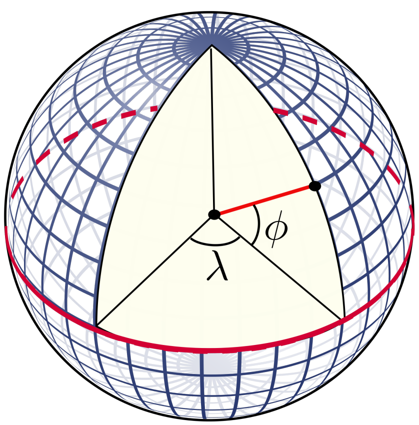
What does “spatially referenced” mean?

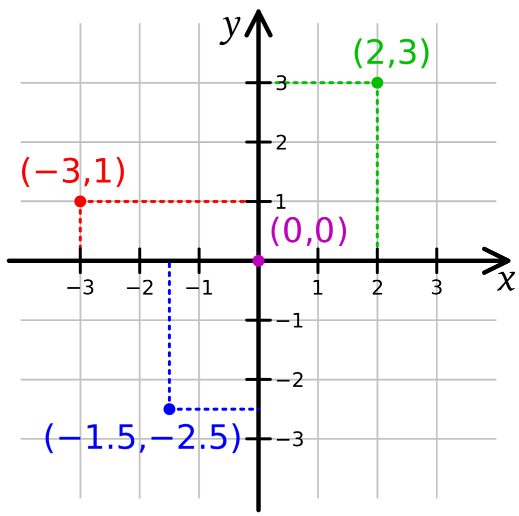
Spatial Coordinates

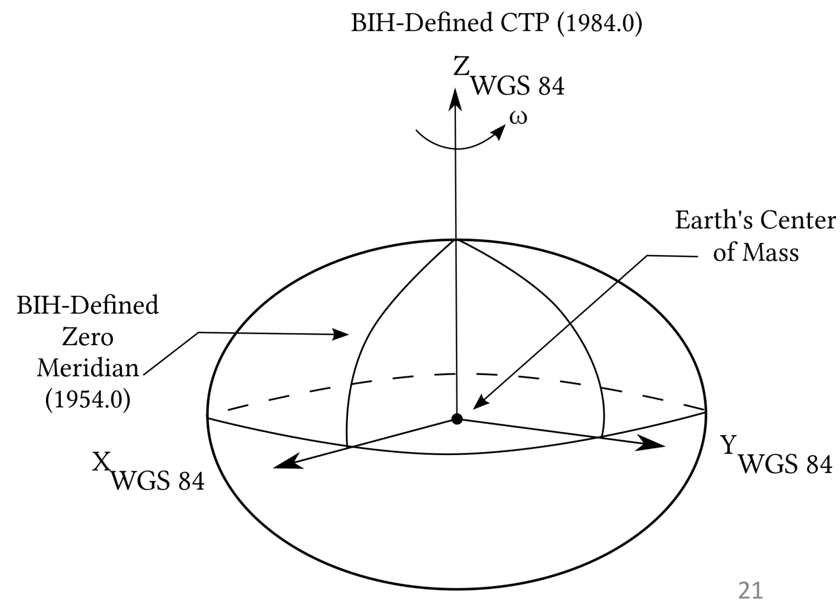
e.g., latitude and longitude; X and Y

Spatial Reference System

e.g., WGS84; Cartesian System



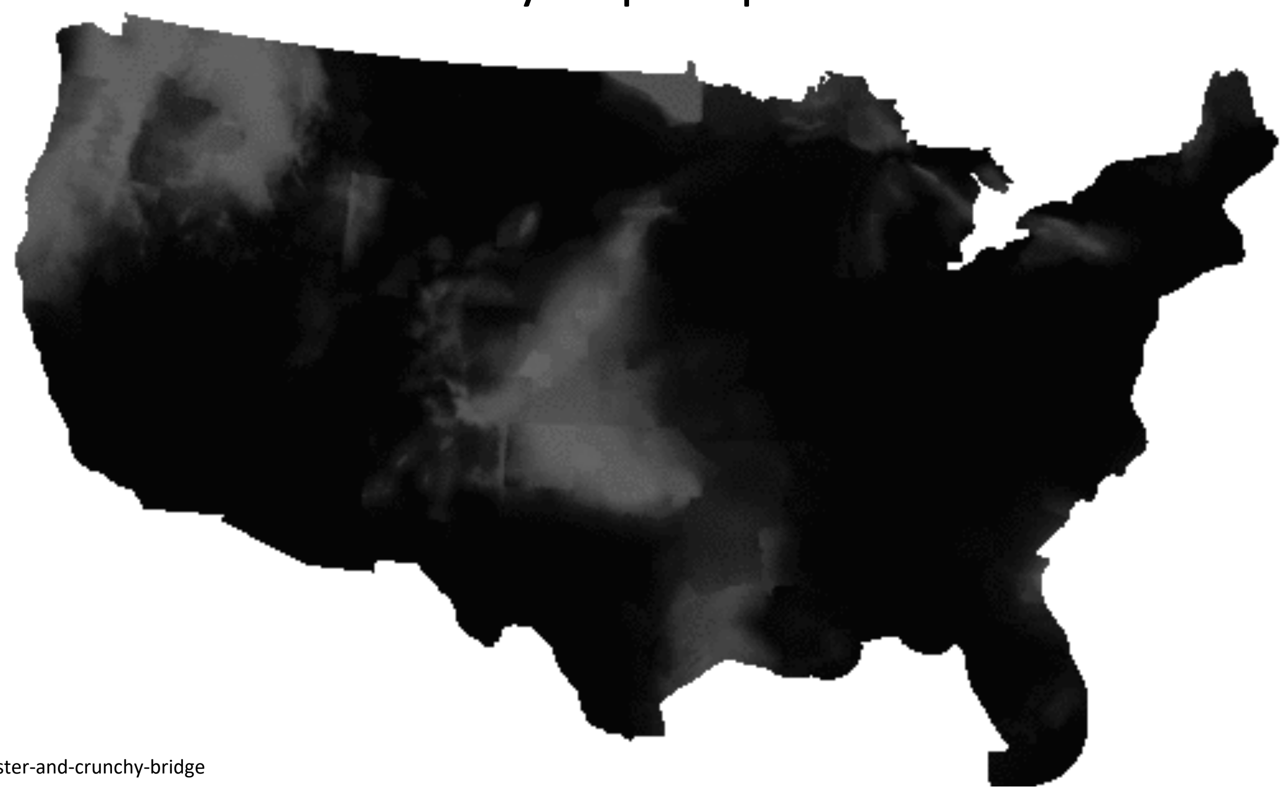




https://en.wikipedia.org/wiki/Geographic_coordinate_system#/media/File:Latitude_and_longitude_graticule_on_a_sphere.svg
https://en.wikipedia.org/wiki/Cartesian_coordinate_system#/media/File:Cartesian-coordinate-system.svg
https://en.wikipedia.org/wiki/World_Geodetic_System#/media/File:WGS_84_reference_frame_(vector_graphic).svg

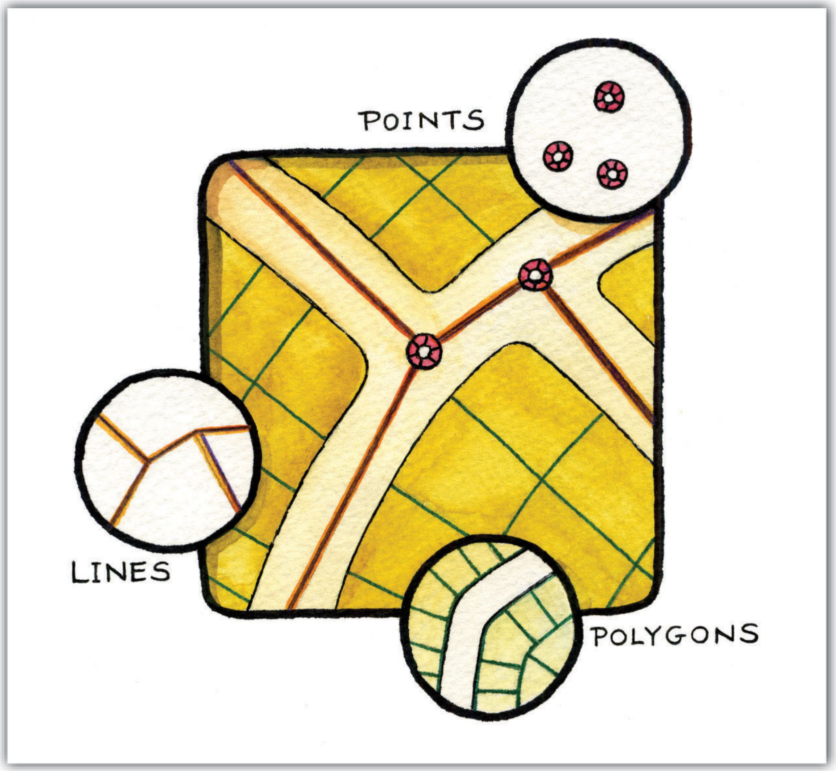
Spatial Data Representations – Raster Data

Probability of precipitation



https://blog.crunchydata.com/blog/postgis-raster-and-crunchy-bridge

Spatial Data Representations – Vector Data



https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-02-vector-data-models.html

Spatial Data Management

- Spatial Databases
    - Support spatial data manipulations using SQL like languages
    - Require a relational database engine
    - e.g., PostGIS (SF-SQL)
- Spatial Big Data Platforms
    - Support highly parallelized spatial data manipulations
    - Require a Big Data processing platform
    - e.g., GeoMESA + Spark (MapReduce)

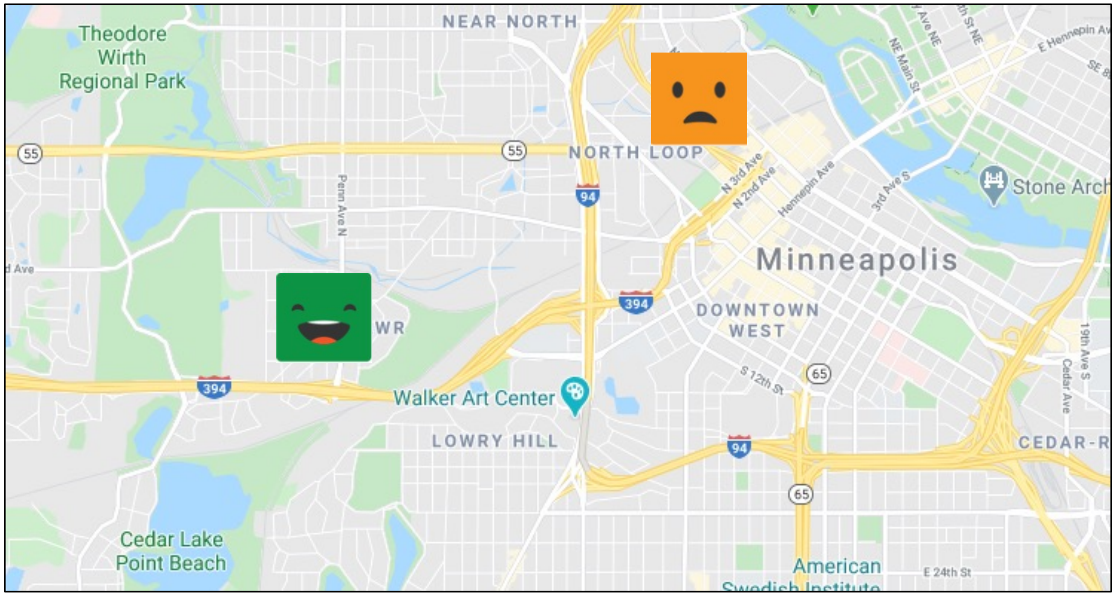
What Make Spatial Data Special?

Spatial Autocorrelation

Nearby houses have similar prices

Spatial Non-stationarity

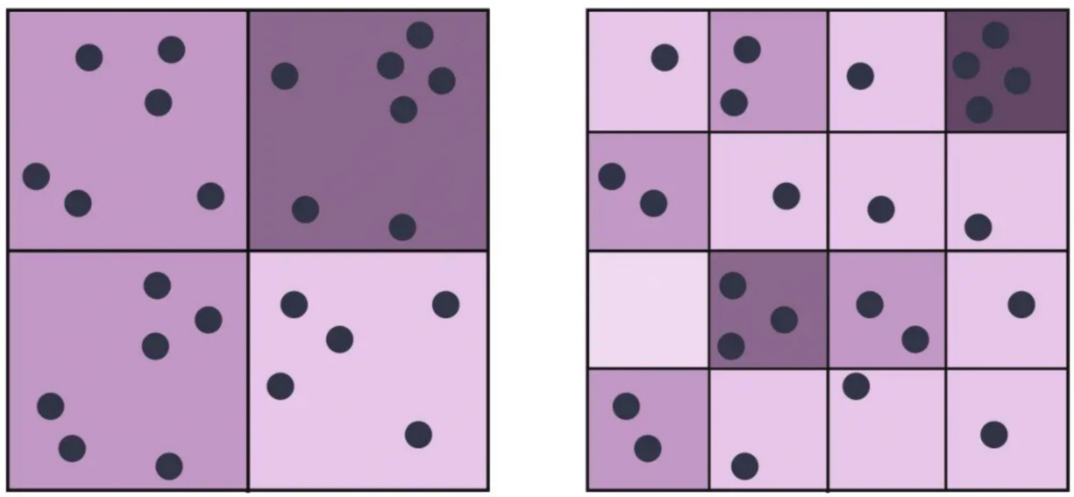
Relationships between variables can change over space



e.g., air quality near highway I-394 can be very different depending on their locations

Modifiable Areal Unit Problem

Scale Effect



https://gisgeography.com/maup-modifiable-areal-unit-problem/

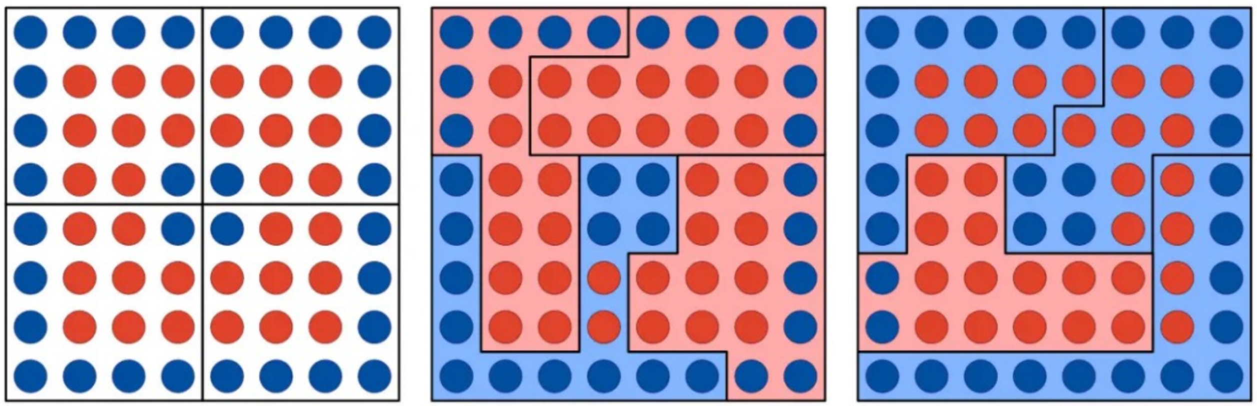
Modifiable Areal Unit Problem

Zonal Effect

https://gisgeography.com/maup-modifiable-areal-unit-problem/

Modifiable Areal Unit Problem

Gerrymandering and Redistricting

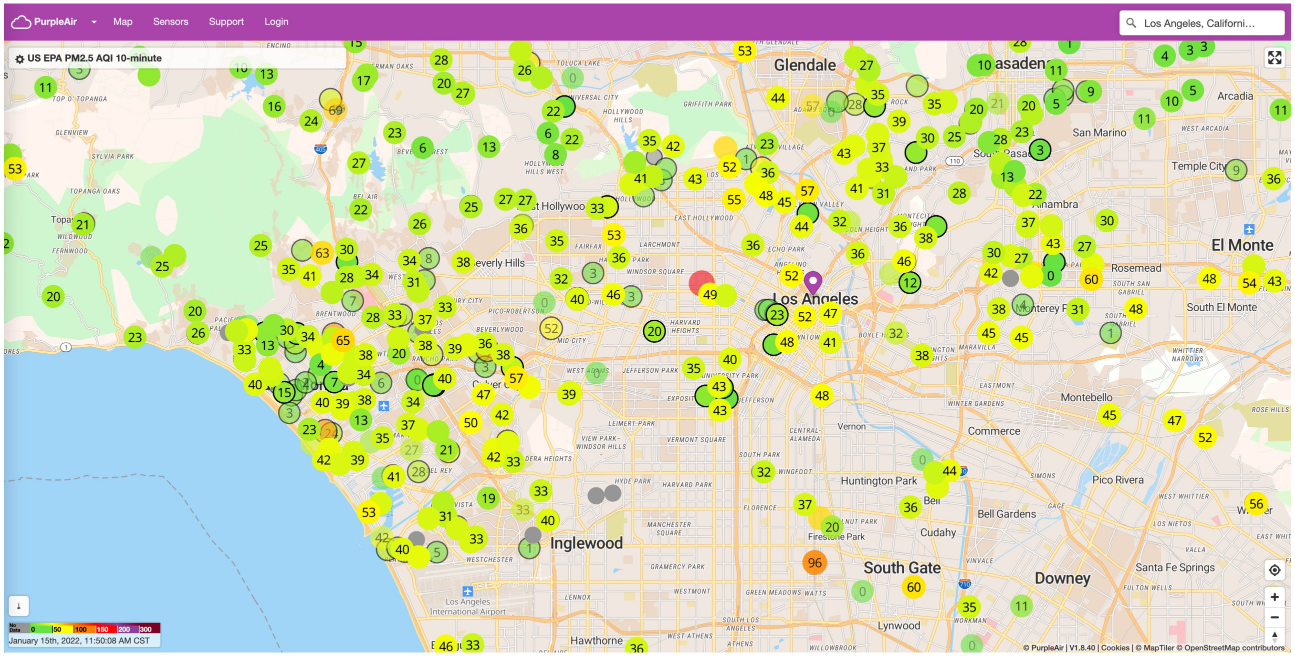


https://gisgeography.com/maup-modifiable-areal-unit-problem/

Spatial Statistics

Geostatistical Data Analysis

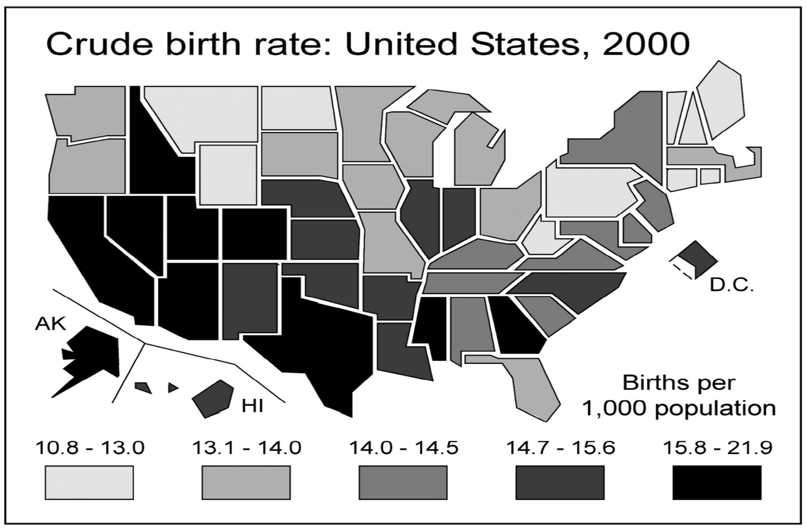
- Data that vary continuously over space, but measured only at discrete locations
- Explore the spatial pattern in the observations
- Quantify the spatial pattern with a function
- Making predictions accounting for spatial structure

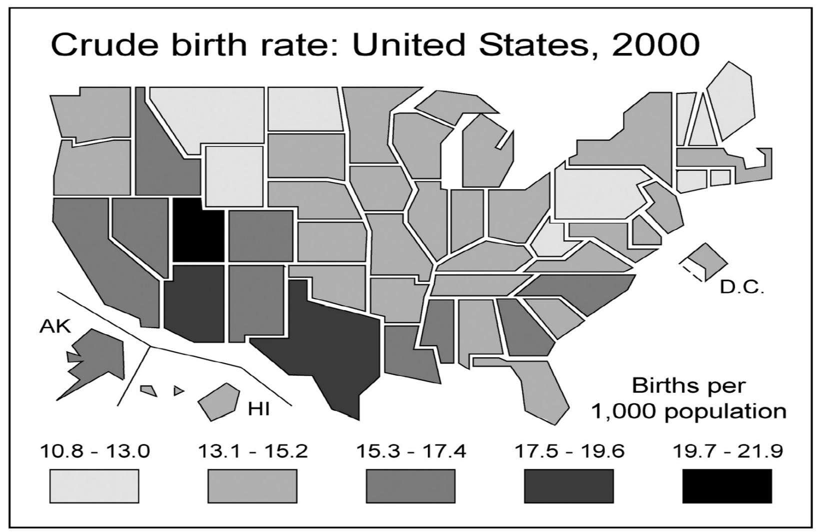


Areal Data Analysis

- Understanding the linkages between areal units
  - e.g., if areas closer to each other are more related, how strong is this patter?

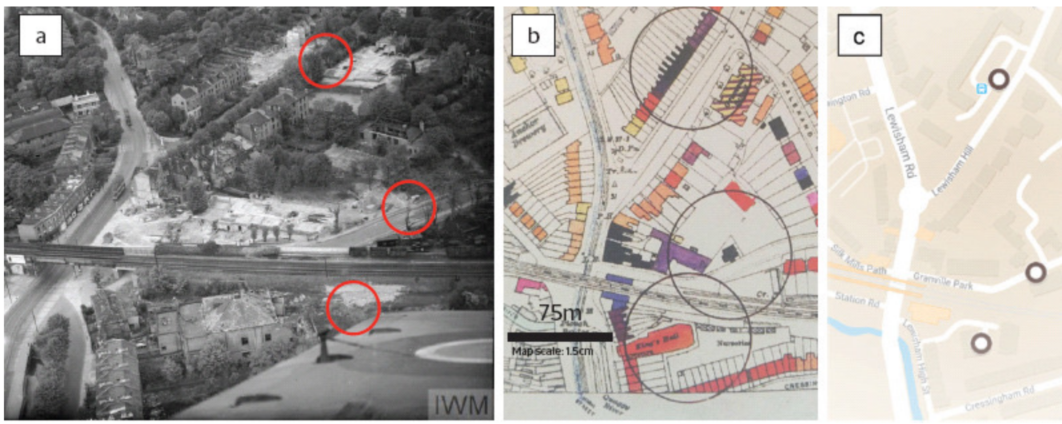
equal-interval vs quantile cut points





Monomier, N. Lying with Maps. Statistical Science 2005, 20(3) 215222.

Point Pattern Analysis



A cluster of three V-1s near a railway line in Lewisham, south London. Patterns like these motivated Clarke to test whether their distribution was random. (a) A contemporary aerial photograph of the V-1 sites, showing the large radius of destruction of houses around the bomb sites (© IWM, catalogue no. CH 15109). (b) The same region from the LCC bomb damage maps4 (p. 162), showing the three V-1 hits (black circles). Damage to houses is coded by colour, with black indicating total destruction. (c) The same region in the Google Maps layer (hit locations V1.509–V1.511).

https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2019.01315.x

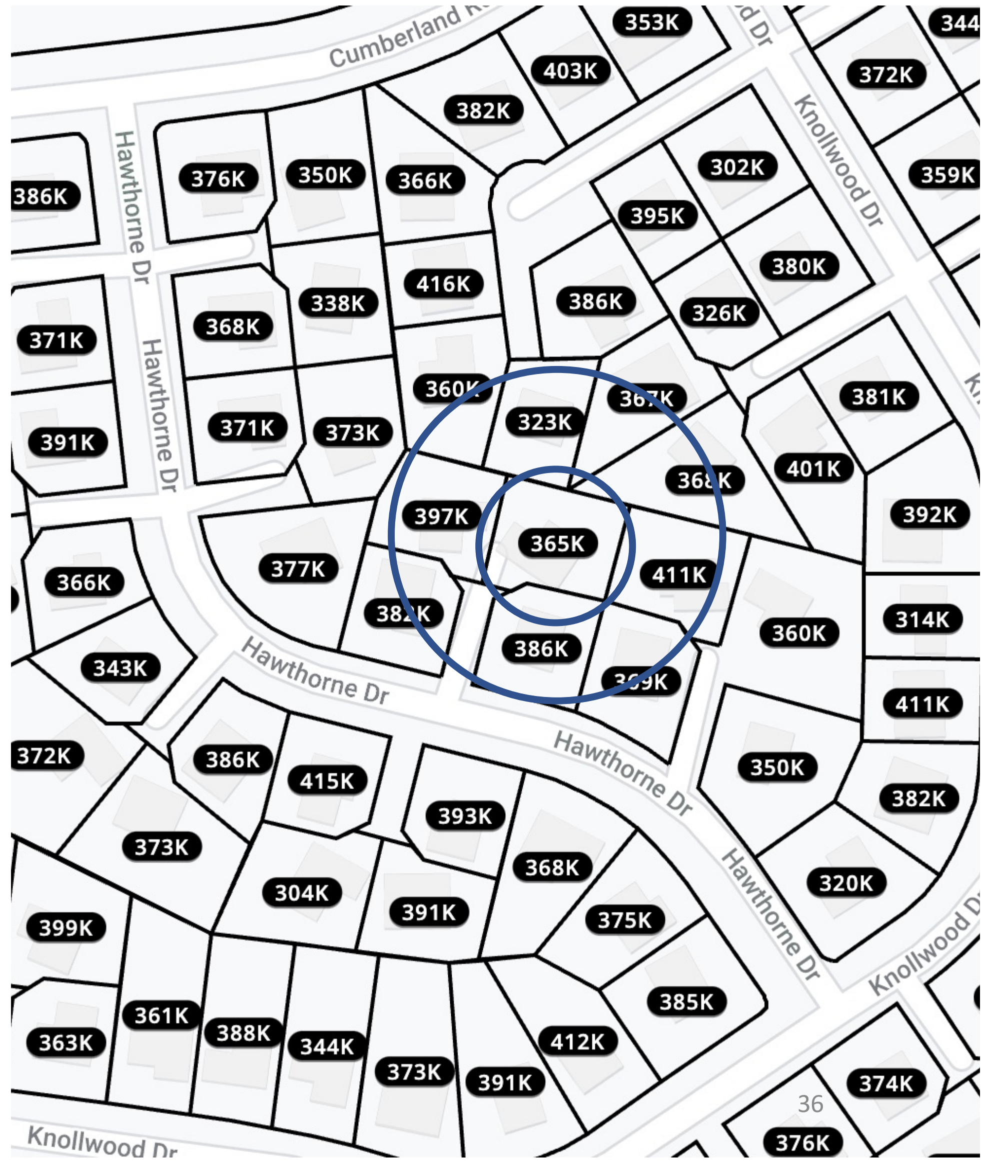
- Is there a regular or clustering pattern in the points?
- Are points closer together than they would be by chance?
- Are the points more regularly spaced than they would be by chance?
- Can we define a point process that our events follow?

Spatial Data Analytic Example

Housing price estimate

K-Nearest Neighbors

- 9 immediate neighbors
- (360+323+367+368+411+369+386+382+397)/9 = 373.66



Inverse Distance Weighting

Point observations:

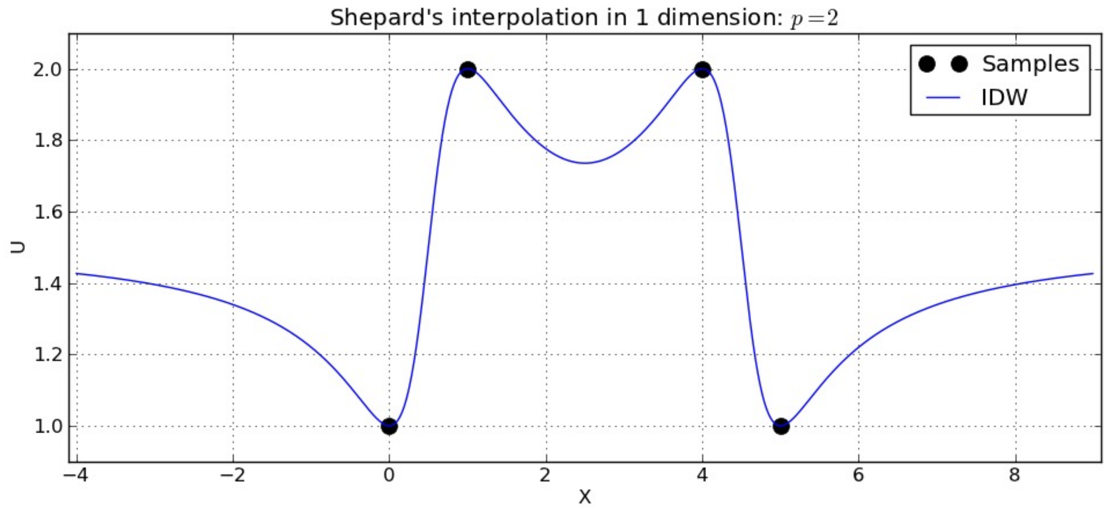
$$u(x): x \to \mathbb{R}, \quad x \in \mathbf{D} \subset \mathbb{R}^n,$$

$$[(x_1, u_1), (x_2, u_2), \ldots, (x_N, u_N)].$$

Interpolated value at X:

$$u(\mathbf{x}) = \begin{cases} \dfrac{\sum_{i=1}^{N} w_i(\mathbf{x}) u_i}{\sum_{i=1}^{N} w_i(\mathbf{x})}, & \text{if } d(\mathbf{x}, \mathbf{x}_i) \neq 0 \text{ for all } i, \\ u_i, & \text{if } d(\mathbf{x}, \mathbf{x}_i) = 0 \text{ for some } i, \end{cases}$$

$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p}$$



https://en.wikipedia.org/wiki/Inverse_distance_weighting

KNN & IDW

- IDW, p=2

$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p}$$

Weight = 1 for immediate neighbors

360+323+367+368+411+369+386+382+397

Weights = 1/4 for two step neighbors

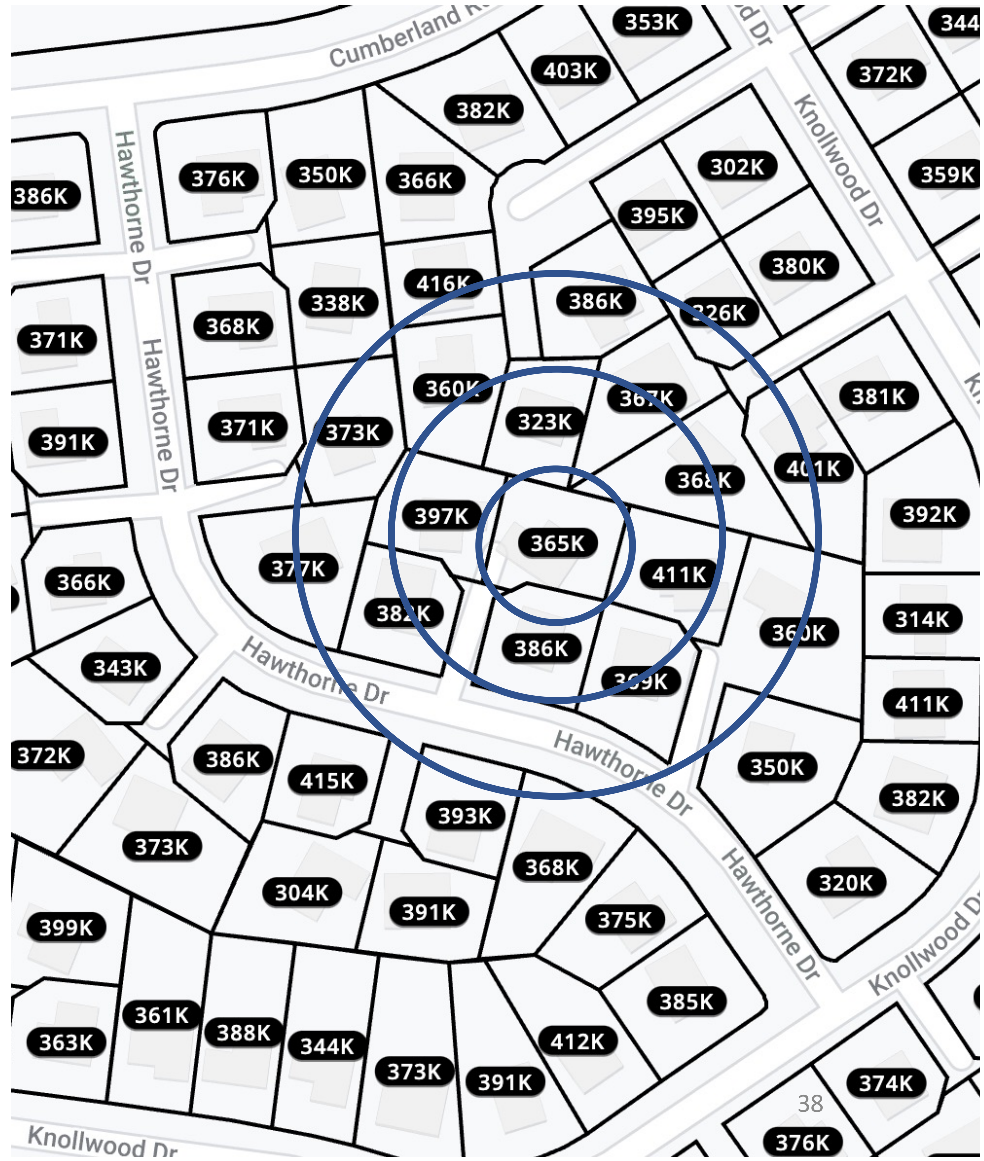
1/4(401+326+386+416+338+373+377+415+393+368+350+360)

Sum of Weights = 1*9 + ¼*12 = 12

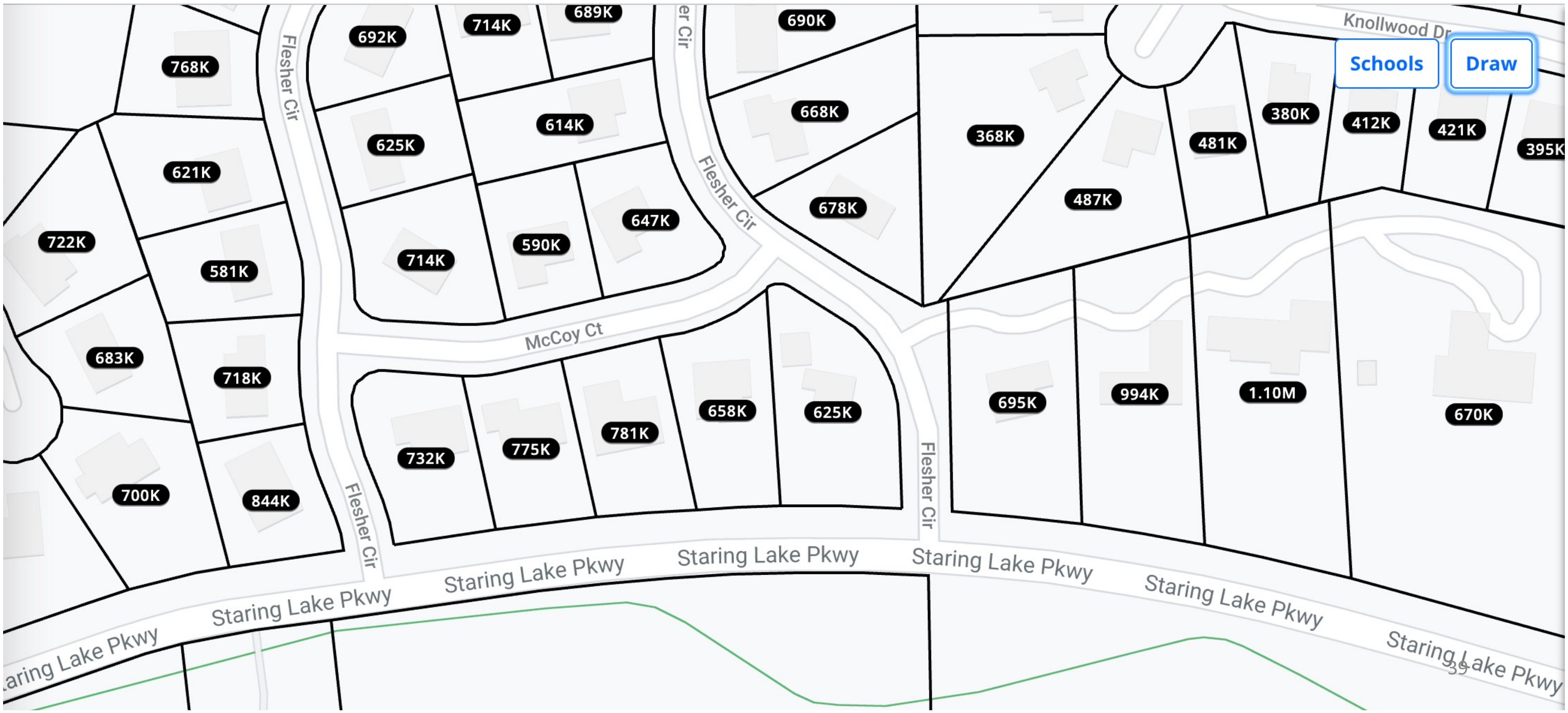
4488.75/12 = 374

- Recall K-Nearest Neighbors

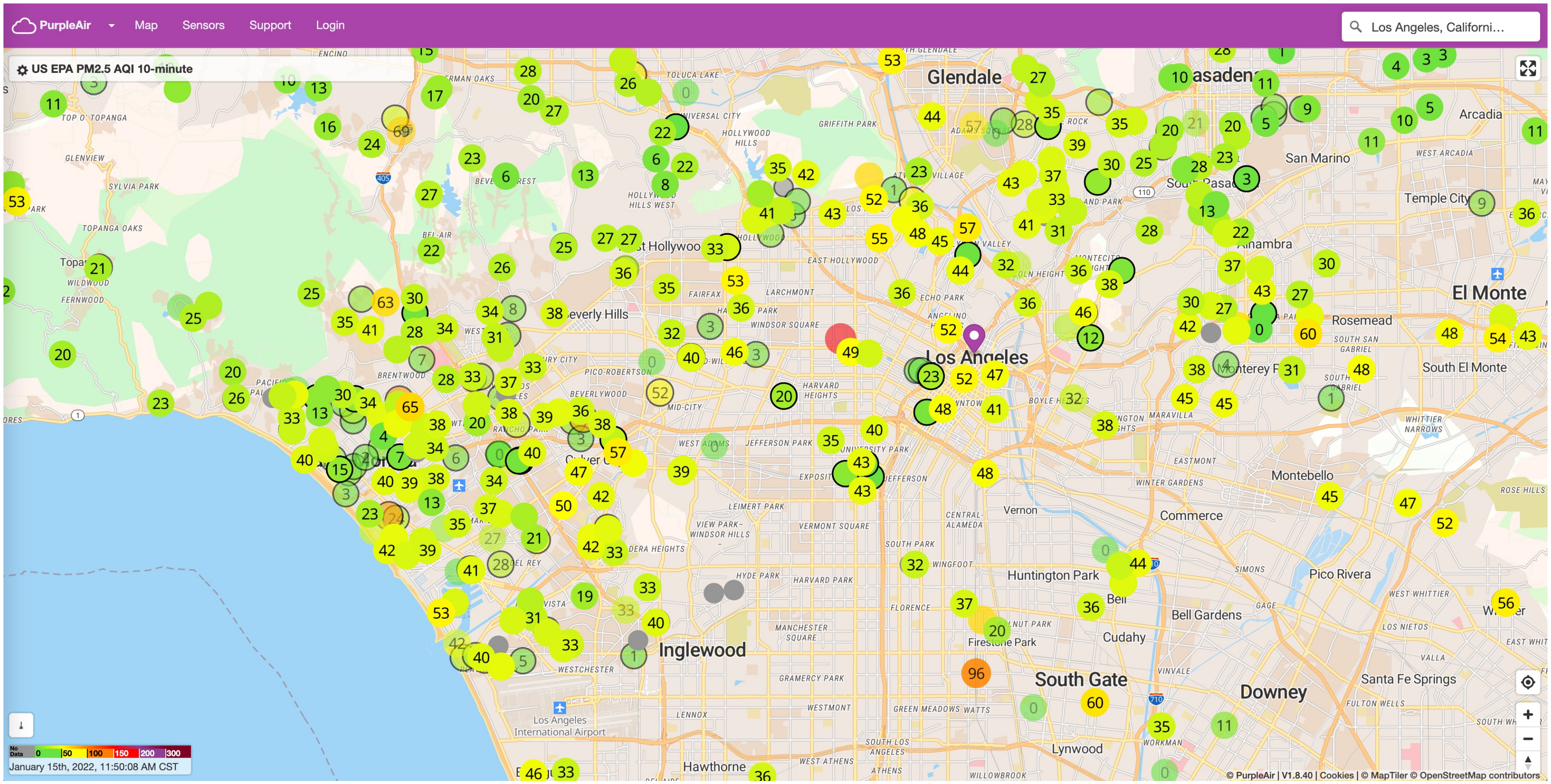
373.66



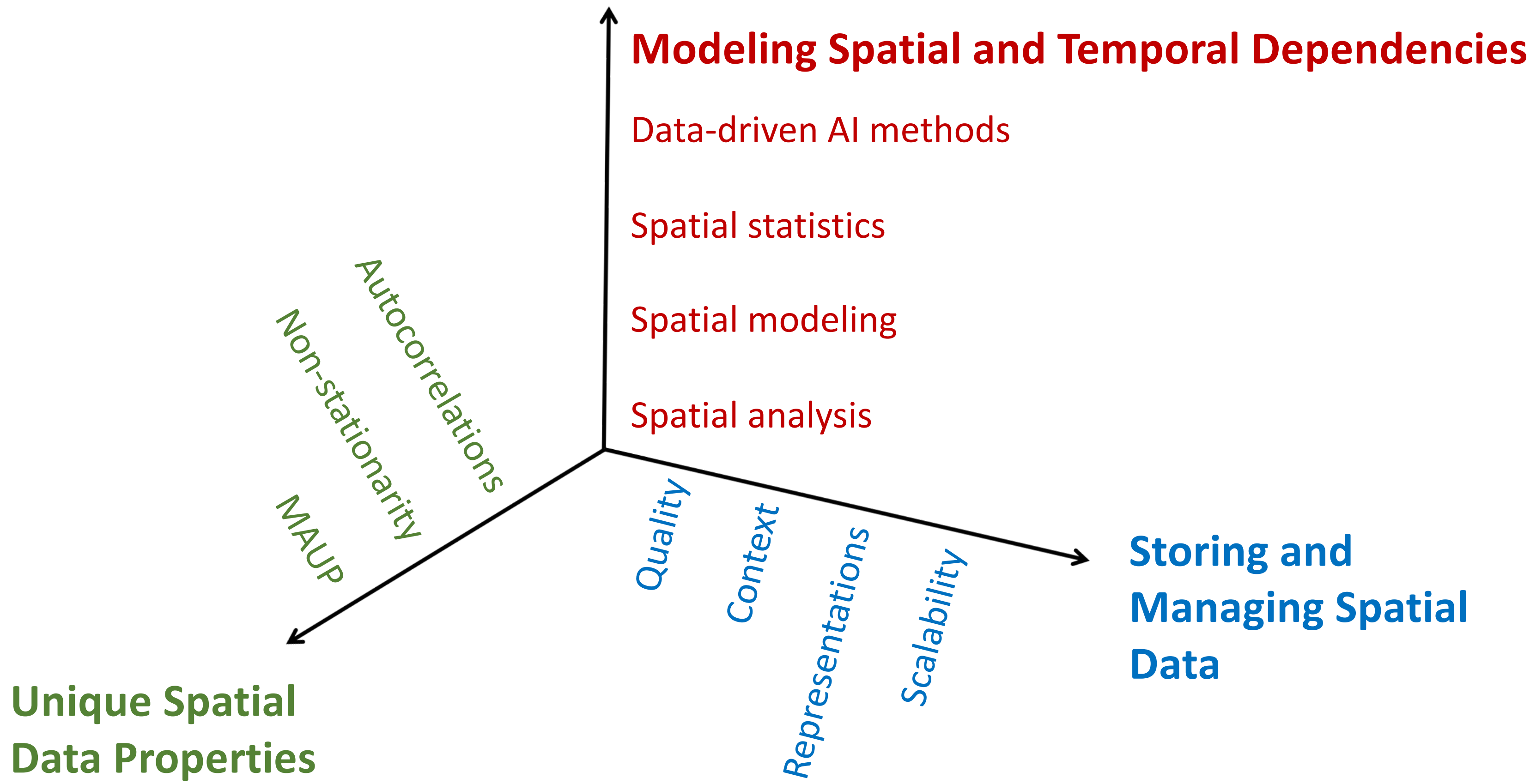
Will either KNN or IDW work here?



Will either KNN or IDW work here?



What Matters When Dealing with Spatial Data?



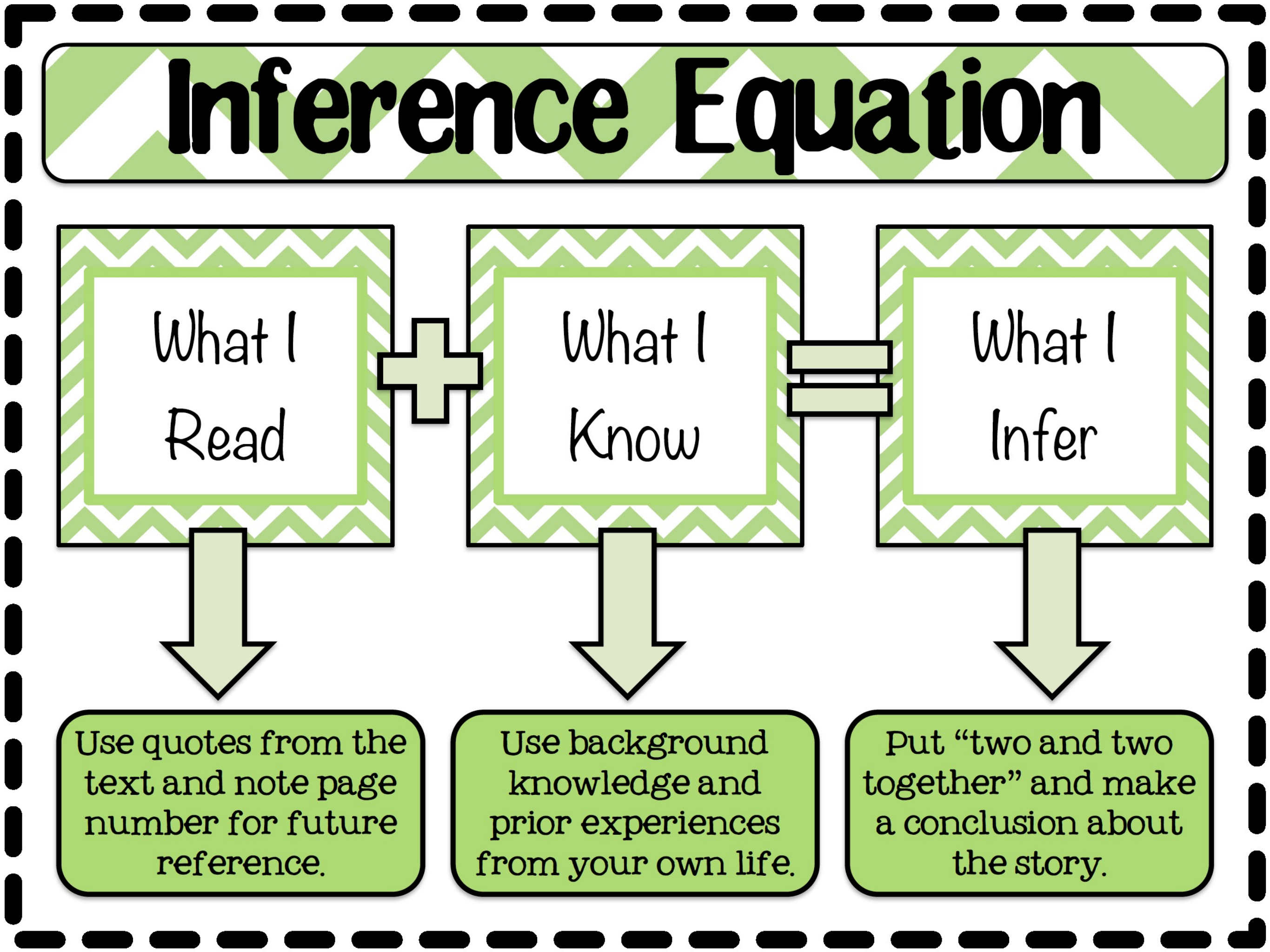
This Course

Course Themes

- Explore ways to store and manage spatial data, including
  - Spatial databases
  - Big Data platforms
  - (if we have time) Ontology and Knowledge Graph
- Look into how deep learning & data mining technologies solve real-world problems utilizing the unique spatial data properties, including topics in
  - Location time-series data mining (e.g., air quality prediction and trajectory mining)
  - Computer vision (e.g., object detection from overhead imagery)
  - (if we have time) Natural language processing (e.g., toponym detection from documents),

Smarter AI

Data-driven technologies that can incorporate prior knowledge derived from spatial data



https://www.teacherspayteachers.com/Product/Making-Inferences-Inference-Equation-Poster?utm_source=sendgrid&utm_medium=email&utm_campaign=post_purchase_confirmation_control&utm_content=01%2F12%2F2022

Prerequisites

- Programming language
  - Python
- Basic understanding of machine learning
  - (We will) Focus on deep learning with PyTorch
- Basic understanding of databases
  - Some familiarity with SQL

Course Tools

- We will provide some background of these tools, but it will be fast paced.
  - Postgres + PostGIS
  - GeoMESA + Spark
  - PyTorch

(Your) Course Work

- Five Homework Assignments (50%, 10% each)
  - No regrading
  - One-week late penalty – 20%; 0 points after one week
  - Free five-day extensions
    - You can use these five days on homework however you want until the last day of the class
    - No more extension days will be given for any reason
    - You need to let your TA know if you are using free days when you submit your homework
- Weekly Quizzes (30%)
- Final project (20%)

Homework Assignments (tentative)

- Using GeoMesa + Spark for efficient spatial data join & aggregation
- Air quality prediction using time-series clustering and random forest
- Road extraction from satellite imagery using deep learning models and existing contextual data
- Air quality prediction using deep learning models and spatial heuristics
- Car detection from satellite imagery using deep learning models and prior knowledge

Final Project Guidelines

- 1 or 2 people, with the following deliverables:
  - Proposal presentation (submit slides for grading)
  - Final project presentation (submit slides for grading)
  - Final report (4 page maximum)
- MS/Senior Undergrad Students
  A comparison of selected state-of-the-art methods for solving a spatial AI problem (e.g., object detection from satellite imagery)
- MS/PhD Students
  Develop a complete research work, which could be related to your research direction

Presentation Guidelines

- Project proposal 10 mins
- Final project presentation 15 mins
- Your presentations need to address the following questions:
  - “What is the project trying to do?”,
  - “How is it done today, and what are the limits of current practice?”,
  - “What is your approach, and what is new in your approach?”,
  - “Who cares? If you succeed, what difference will it make?”,
  - “How do you know if your approach is successful?”, and
  - “What are the future extensions?”

  This is the modified version of the famous “Heilmeier Catechism”: http://www.darpa.mil/work-with-us/heilmeier-catechism

Course Grades

Grades will range from A through F. The following is the breakdown for grading:

94 – 100 = A	74 – 76 = C
90 – 93 = A-	70 –73 = C-
87 – 89 = B+	67 – 69 = D+
84 – 86 = B	64 – 66 = D
80 – 83 = B-	60 – 63 = D-
77 – 79 =C+	Below 60 is an F

Course Staff

- TAs:
  - Yijun Lin, lin00786@umn.edu
  - Zekun Li, li002666@umn.edu
- Office Hours:
  - Instructor: Wednesdays after class
  - TAs: See Piazza for TA office hours and locations

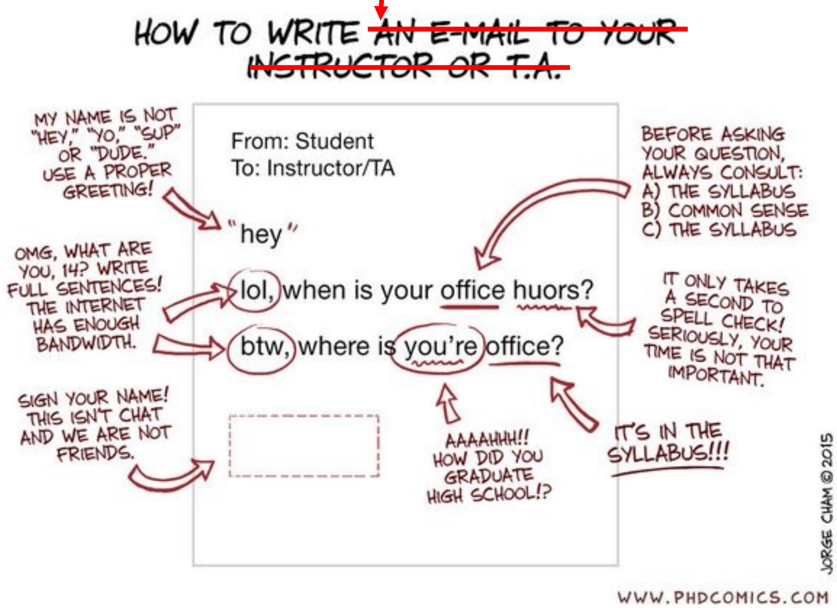
Course Logistics

- Course websites:
  - https://yaoyichi.github.io/spatial-ai.html
    - Material distributions: e.g., lecture slides (same day after the lecture)
  - https://piazza.com/class/ky91axj4suo9v
    - Discussions
  - Canvas
    - Grading, assignment submission, etc.

Communications

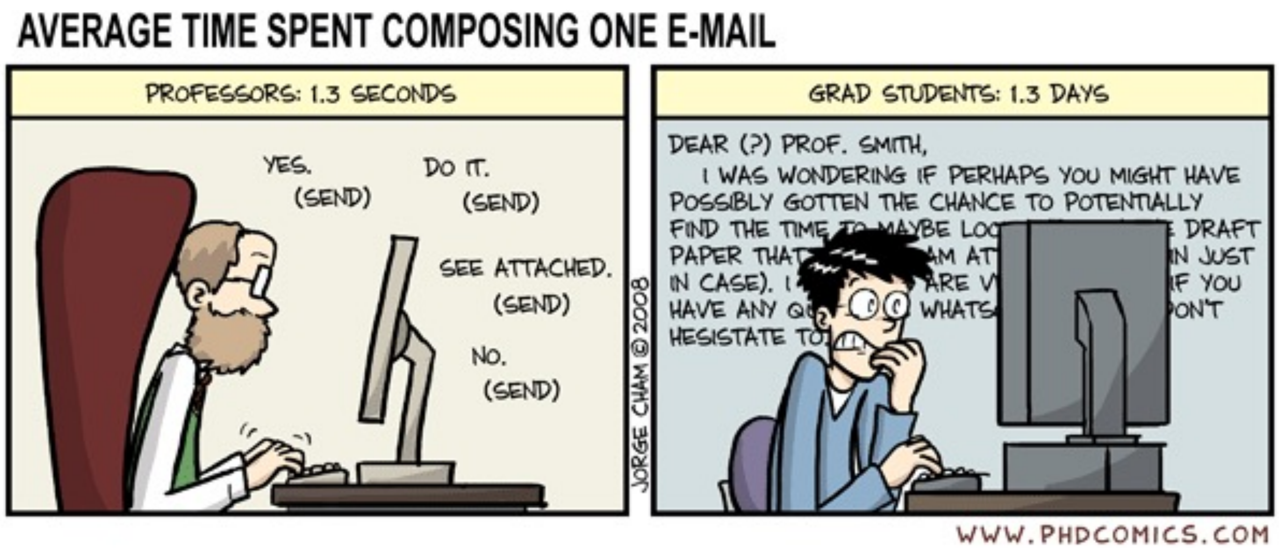
- Discussion board on Piazza:
  - Use the discussion board for all questions and public communication with the course staff

a post on the discussion board



Please do not email us unless…

- We will post course announcements to Piazza (make sure you check it regularly).
- Emails:
  - Do not use emails unless it's personal!

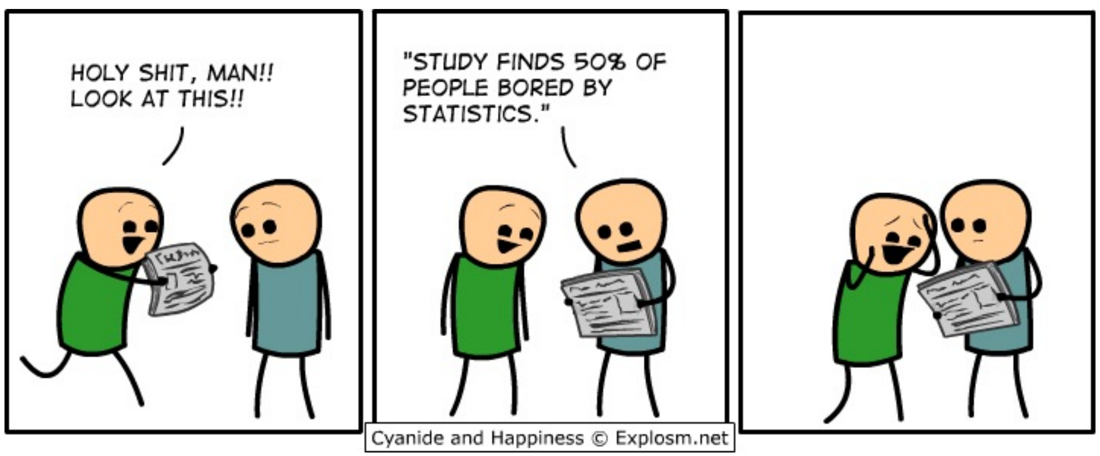


Readings

- Jure Leskovec, Anand Rajaraman, Jeff Ullman, *Mining of Massive Datasets*
  - Cambridge University Press, 2012
  - Available free at: http://www.mmds.org/
- Shashi Shekhar and Sanjay Chawla, *Spatial Databases: A Tour*
  - Prentice Hall, 2003 (ISBN 013-017480-7)
  - http://www.spatial.cs.umn.edu/Book/
- Ian Goodfellow and Yoshua Bengio and Aaron Courville, *Deep Learning*
  - MIT Press, 2016
  - Available free at: https://www.deeplearningbook.org/
- Additional papers

To-Do Items

- Download the online textbook and readings
- Install Spark and Postgres (plus PostGIS) on your machine
- Signup for Piazza



Acknowledgements

- Data Mining
  - Jure Leskovec, Anand Rajaraman, Jeff Ullman
  - Mining of Massive Datasets
  - http://www.mmds.org/
- Spatial Statistics
  - Meredith Franklin
  - Spatial Statistics
  - https://github.com/meredithfranklin/courses/tree/master/Spatial
- Deep Learning
  - © Alexander Amini and Ava Soleimany
  - MIT 6.S191: Introduction to Deep Learning
  - https://introtodeeplearning.com/
- Gil, Yolanda (Ed.) Introduction to Computational Thinking and Data Science. Available from http://www.datascience4all.org