UNIVERSITY OF CALIFORNIA
UC RIVERSIDE

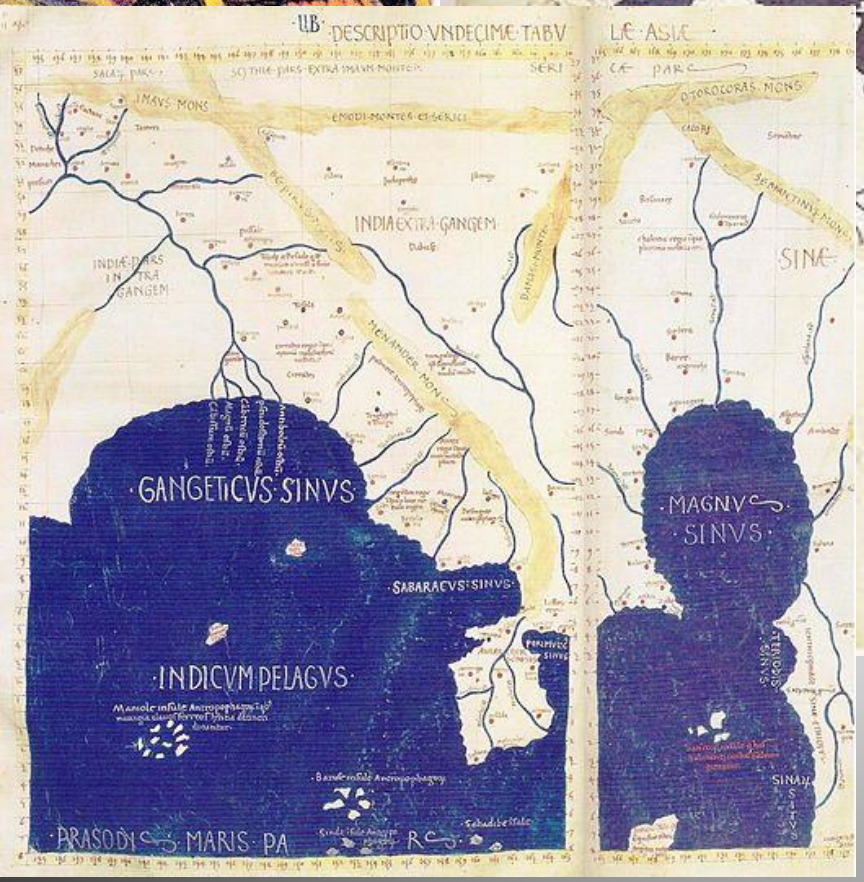
Spatial Big Data Management: A Brief History, State, and Applications

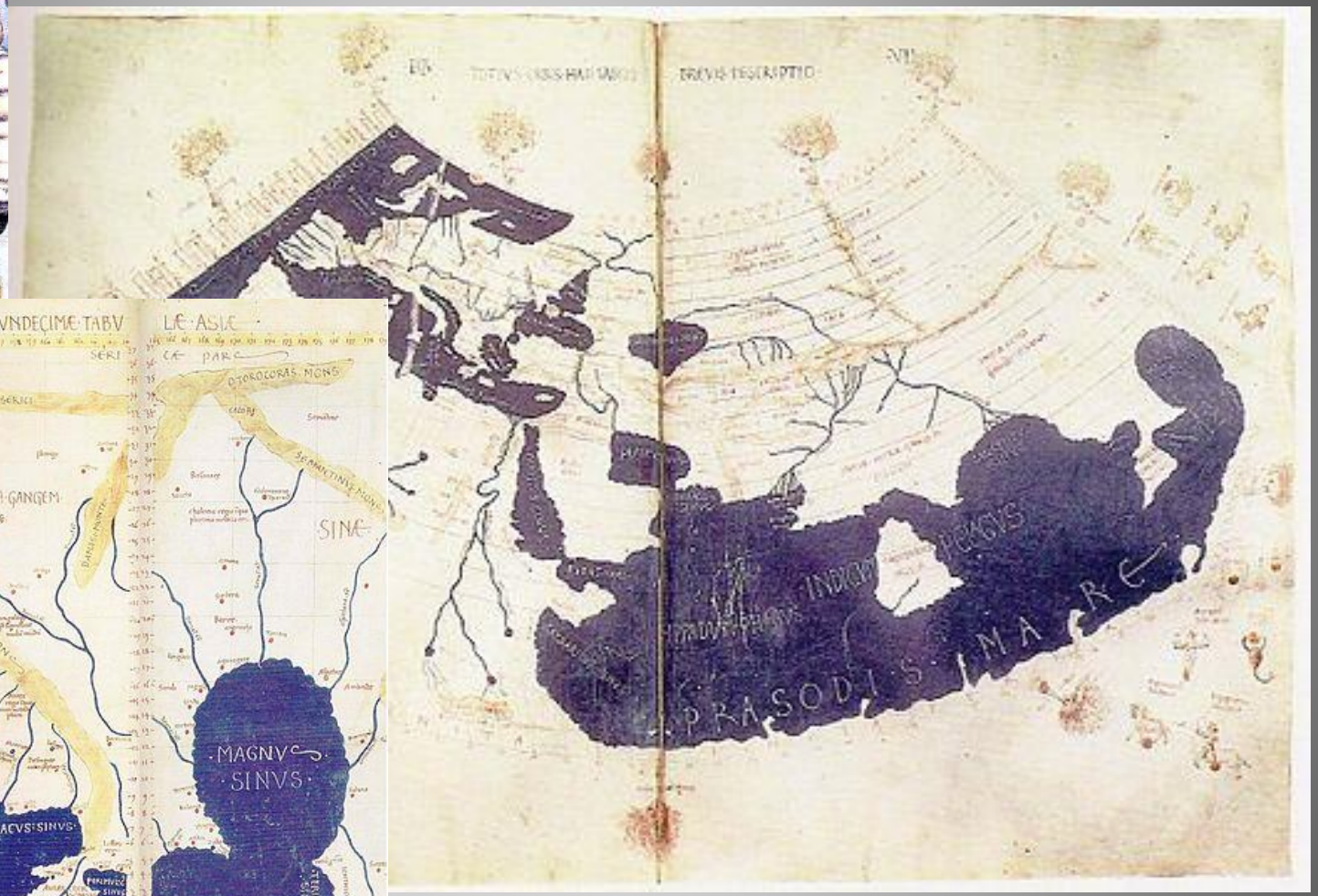
Amr Magdy
Computer Science and Engineering

Email: amr@cs.ucr.edu
www.cs.ucr.edu/~amr/

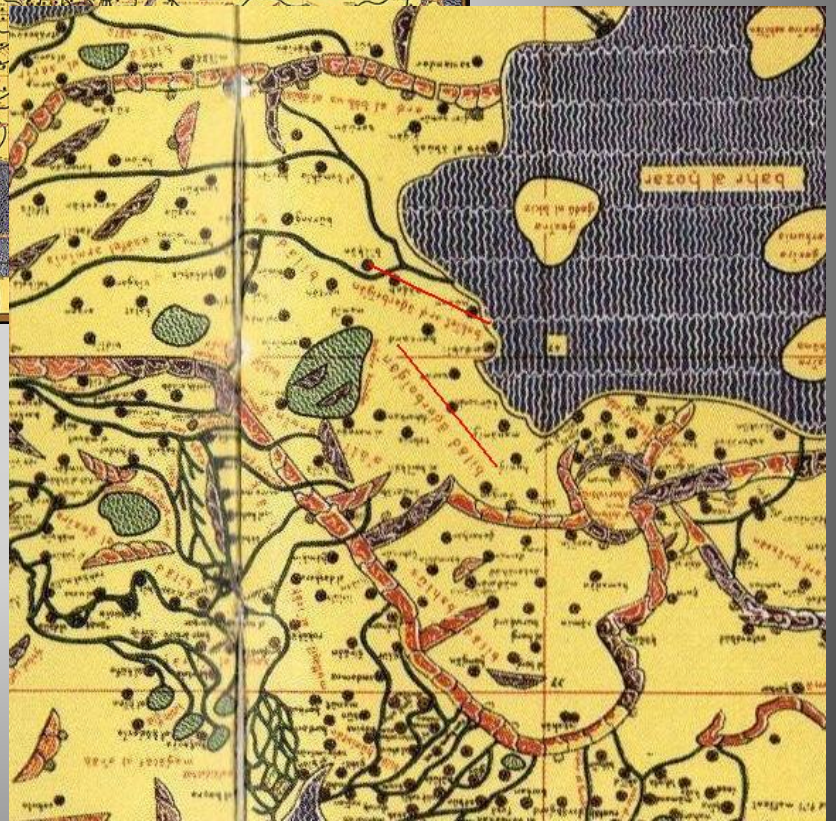

Once upon a
time...

Claudius Ptolemy (AD 90 – AD 168)

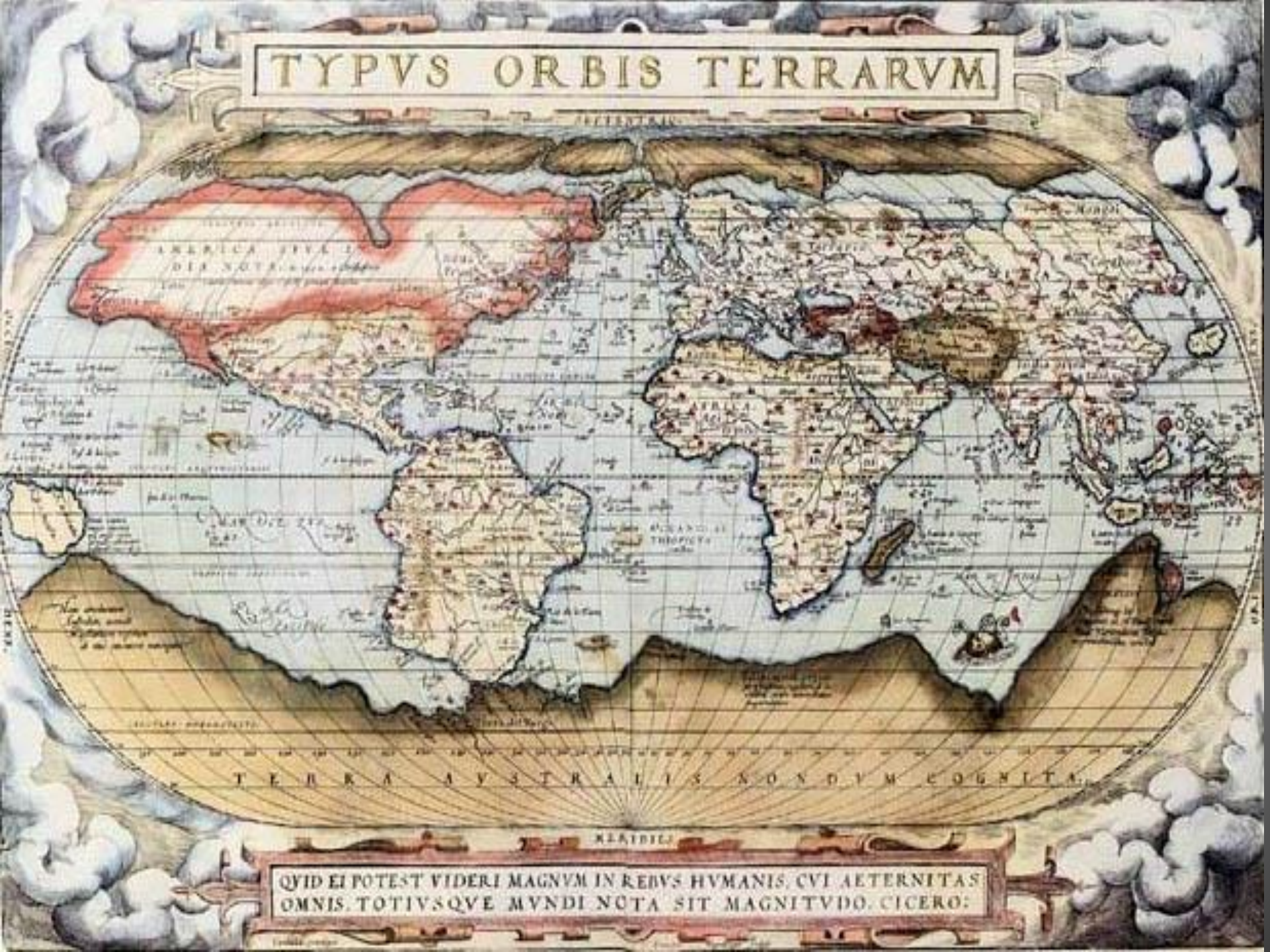




Al Idrisi (1099–1165)

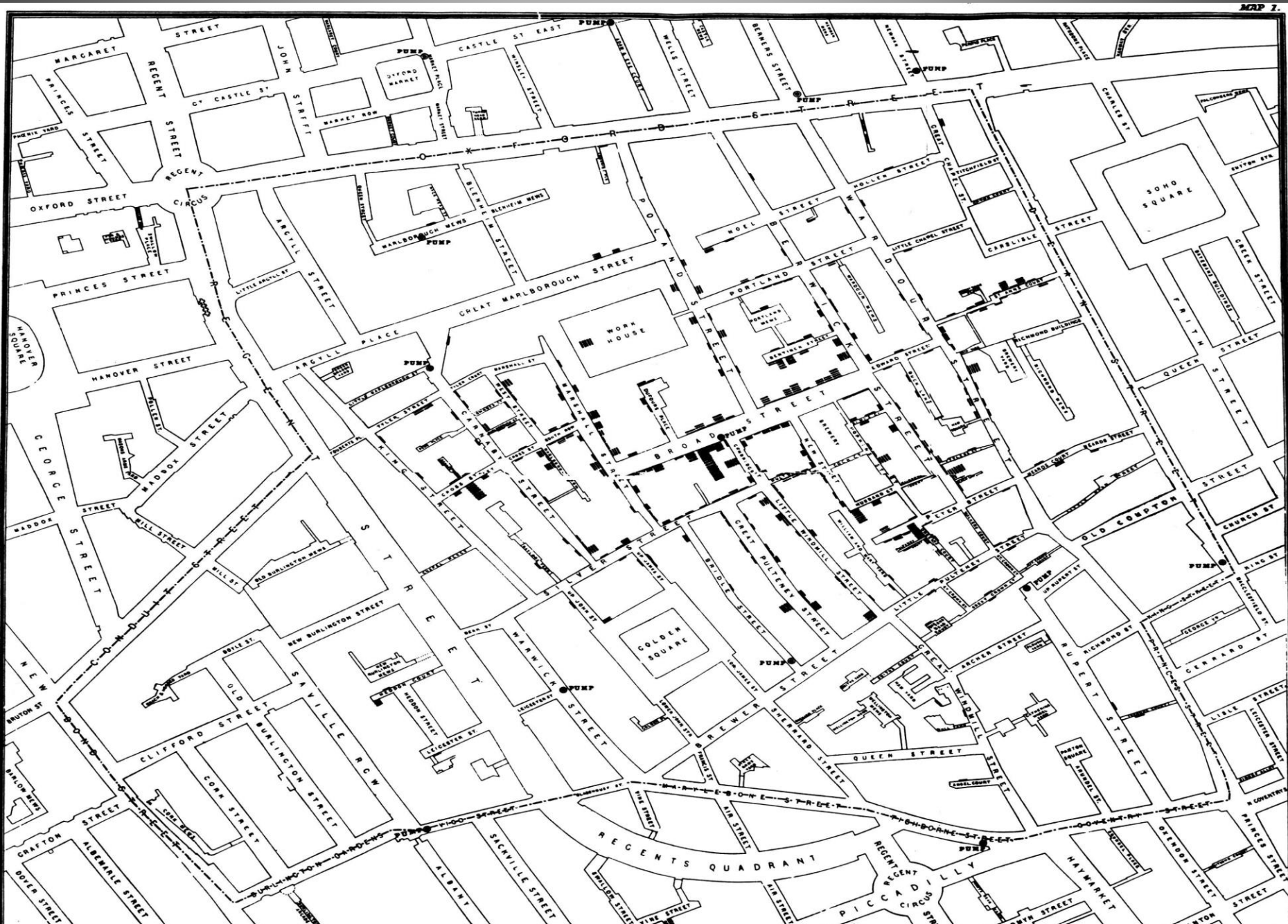


Mar Atlantico
EVROPA
AFRICA
INDIA
Tropicus Cancri
Equinoctialis
Brasil
Mar Indico
Tropicus Capricorni
Mar Pacifico
Mar Meridional

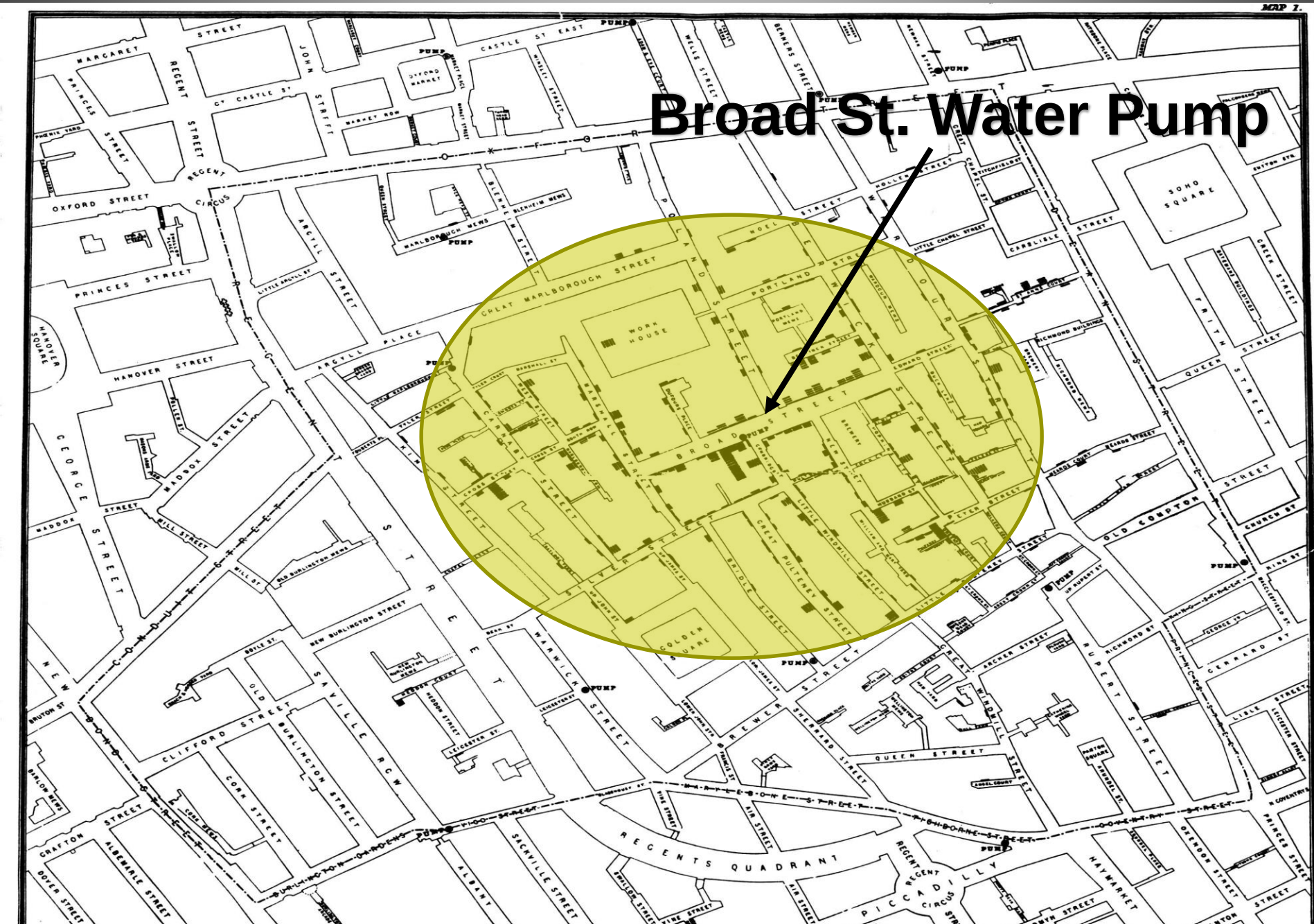

TYPVS ORBIS TERRARVM
TERRA AVSTRALIS NONDVM COGNITA
QVID EI POTEST VIDERI MAGNVM IN REBVS HVMANIS, CVI AETERNITAS
OMNIS, TOTIVSQVE MVNDI NOTA SIT MAGNITVDO. CICERO:


ARGONAVTICA.
EVROPAE PARS.
ASIAE PARS.
LIBYAE PARS.
SAVROMATAE.
SCYTHAE.
GETAE
ARIMASPHAE.
MAEOTIS PALVS.
AXENVM Aequor, Iasone pollutum remige primum.
CAVCASEVM MARE.
CAPPADOCIA.
THRACIA.
CELTAE.
SARDOVM PELAGVS.
TYRRHENVM AQVOR.
IONIVM MARE.
LIBYSTICVM MARE.
PELOPIS REGIO.
Maeotim pelagus
THRACIA
PROPONTIS.
BITHYNIA.
ILLVSTRISSIMO PRINCIPI CAROLO COMITI ARENBERGIO, BARONI SEPTIMONTII, DOMINO MIRVARTII, EQVITI AVREI VELLERIS, ETC.

Cholera cases in the London epidemic of 1854



Cholera cases in the London epidemic of 1854



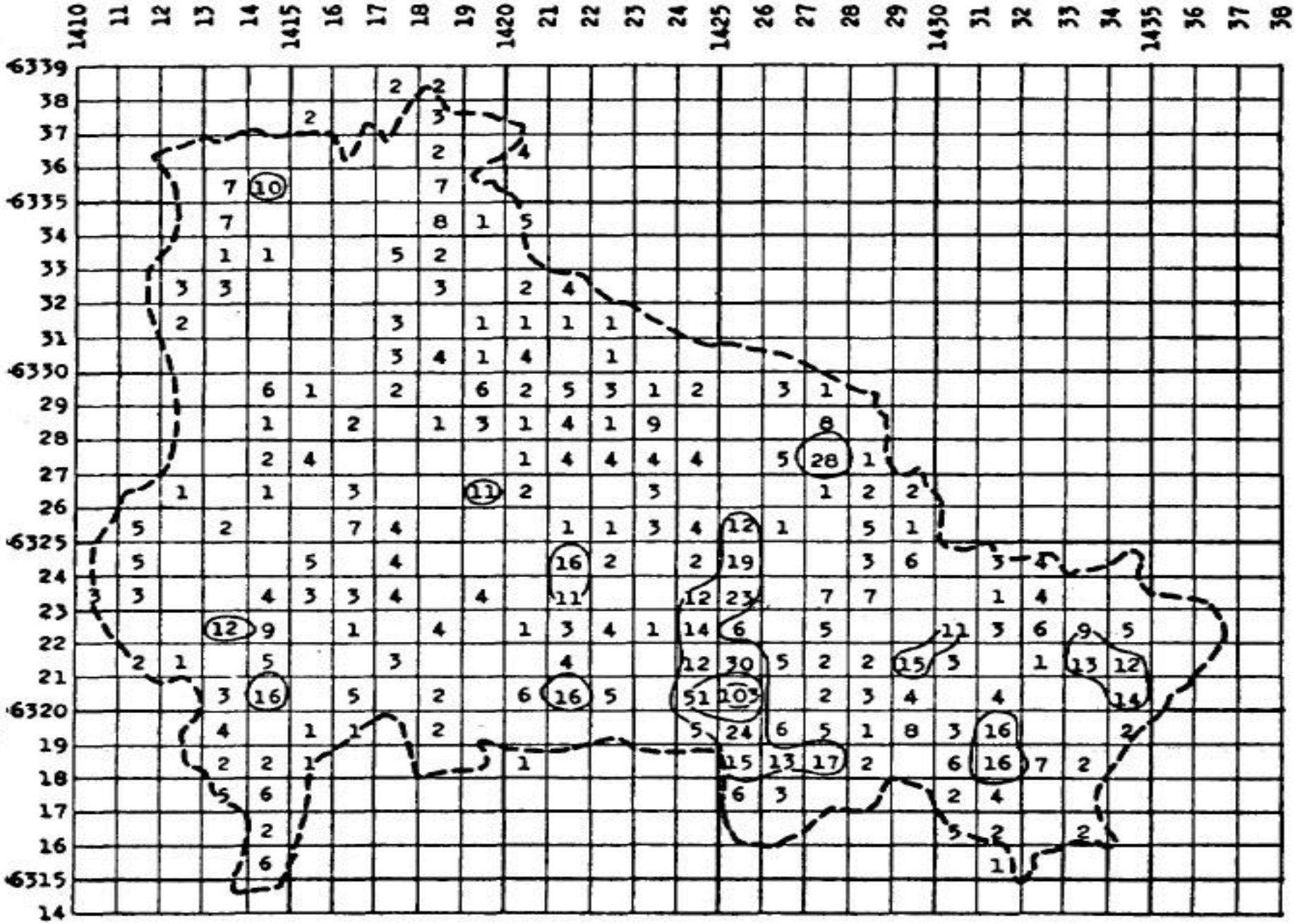


FIGURE 3—Children under 15 years of age in 1940.

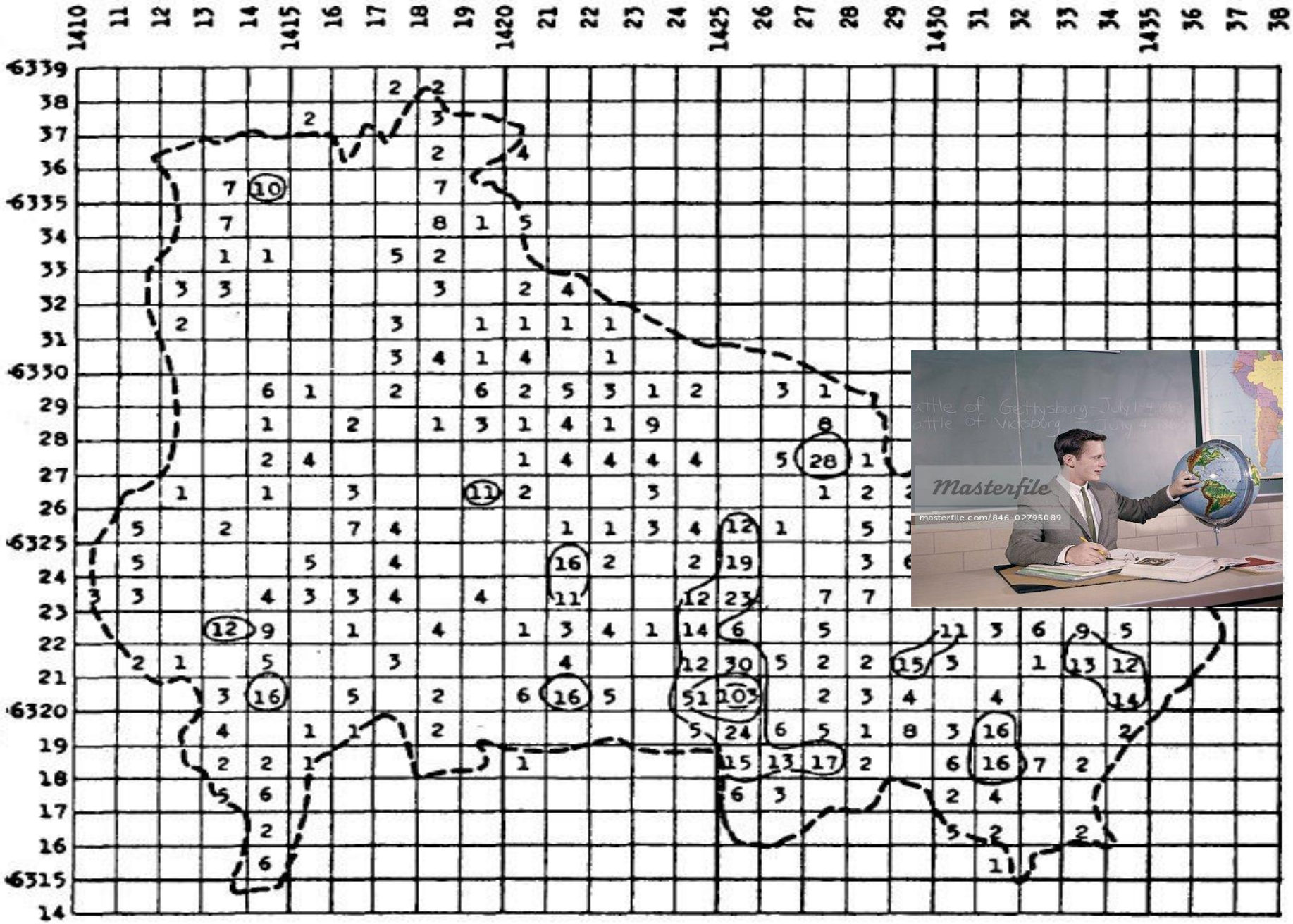
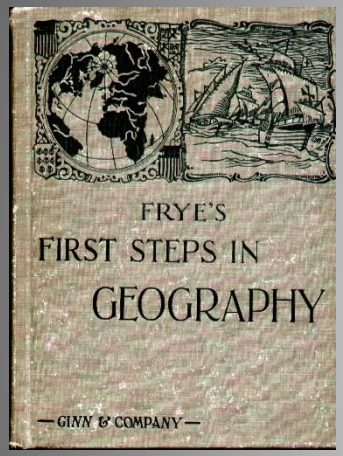


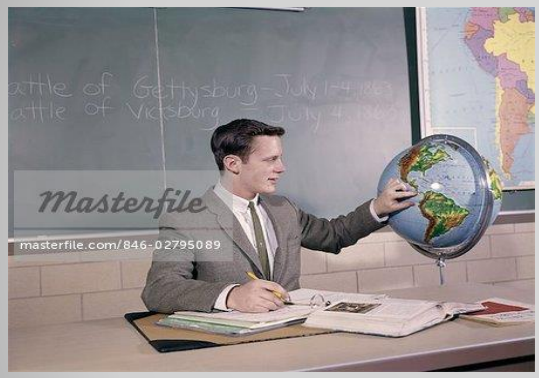
FIGURE 3—Children under 15 years of age in 1940.


UAR
5 M

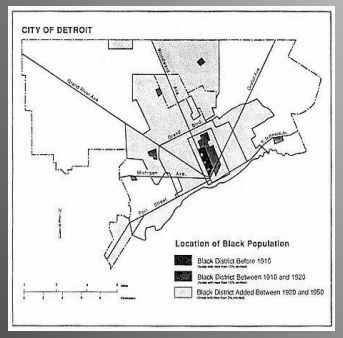

OCÉAN ATLA
Newcastle
Liverpool
Birmingham
Cardiff
Londres
Hambourg
Amsterdam
Rotterdam
Anvers
Le Havre
Paris
Nantes St-Nazaire
Bordeaux
Marseille
Milan
Trieste
Gênes
Madrid
Barcelone
Rome


UNIVAC


FRYE'S
FIRST STEPS IN
GEOGRAPHY
—GINN & COMPANY—


Masterfile
masterfile.com/846-02795089


IBM

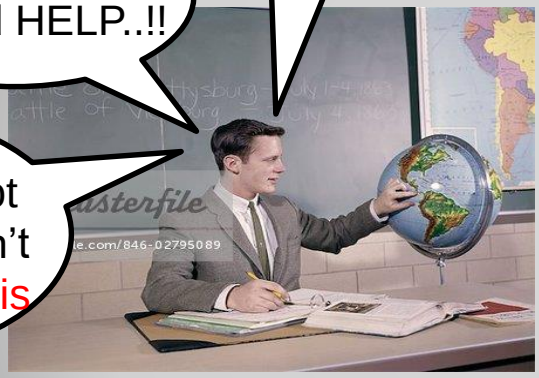

CITY OF DETROIT
Location of Black Population
Black District Before 1910
Black District Between 1910 and 1920
Black District Added Between 1920 and 1950

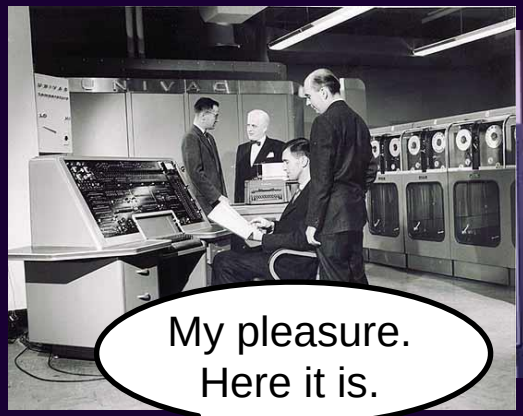

Masterfile
masterfile.com/846-02793618

Cool computer technology..!! Can I use it in my application

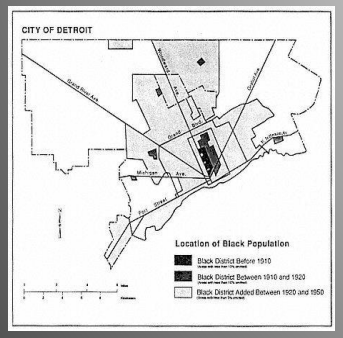
I have **BIG** data. I need HELP..!!

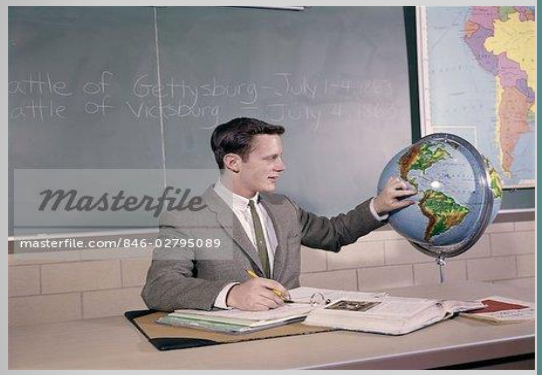
Oh..!! But, it is not made for me. Can't make use of it as is





My pleasure. Here it is.



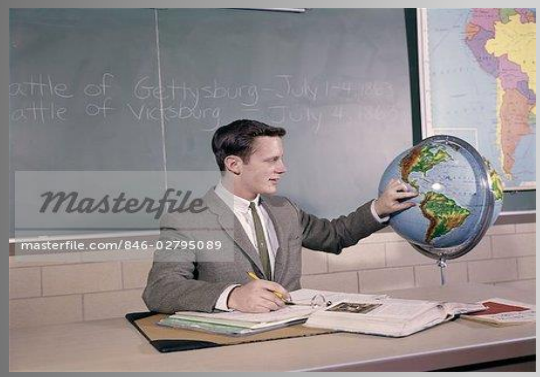

Masterfile
masterfile.com/846-02795089


IBM

1969

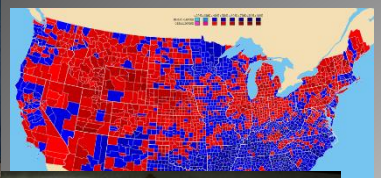
Kindly let me understand your needs

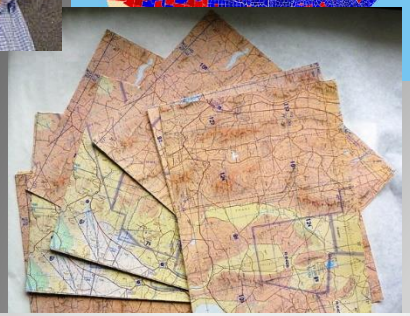
Kindly let me get the technology you have

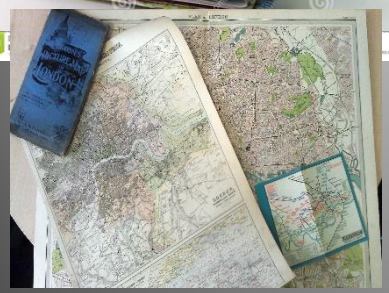


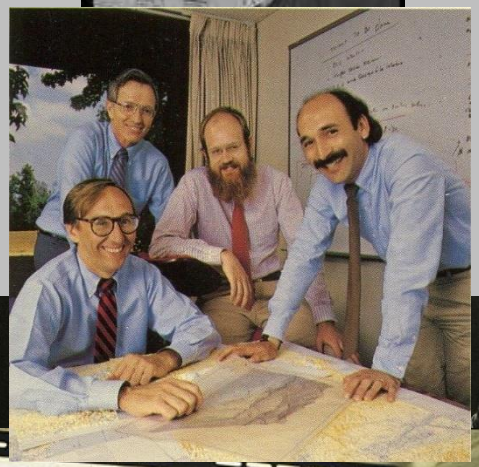
ESRI

ESRI ArcGIS

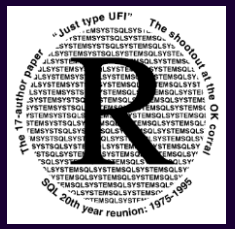









ESRI


"Just type UFI"
The 17-author paper
The shootout at the OK corral
SQL 20th year reunion: 1975-1995


DATABASE
MANAGEMENT
SYSTEMS

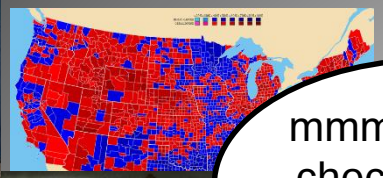

Informix

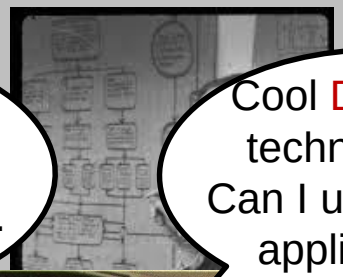

ORACLE
DATABASE
ORACLE

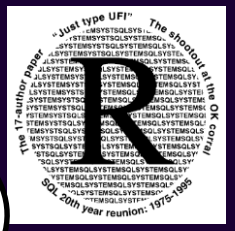
PostgreSQL


AIRLINE BOOKING


Banking







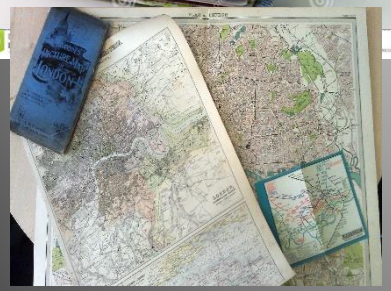
HELP..!! I have **BIG** data. Your technology is not helping me

mmm…Let me check with my good friends there.

Cool Database technology..!! Can I use it in my application?

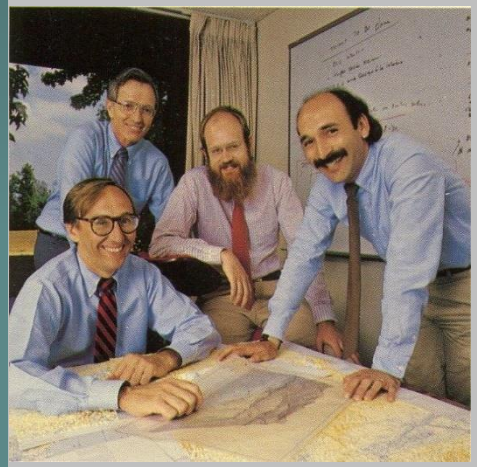
My pleasure. Here it is.

Oh..!! But, it is not made for me. Can't make use of it as is

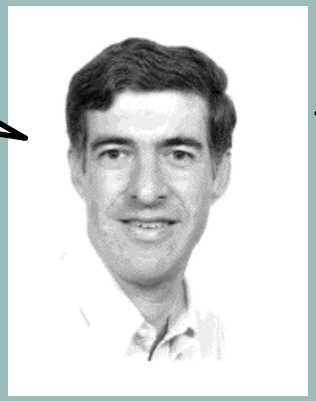


Informix

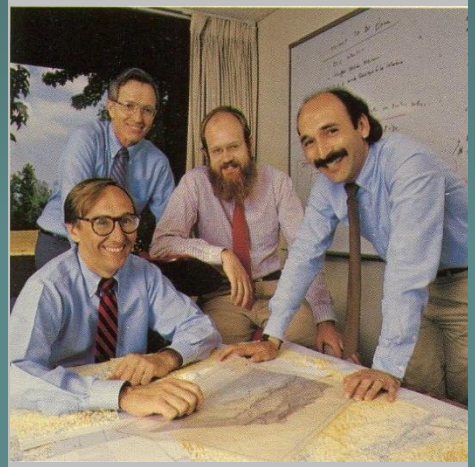
PostgreSQL

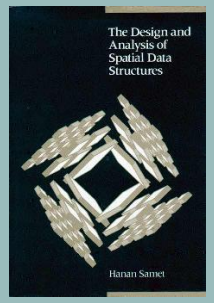


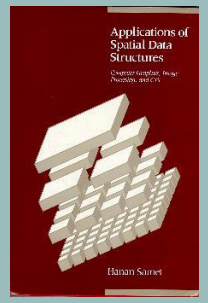
Kindly let me understand your needs

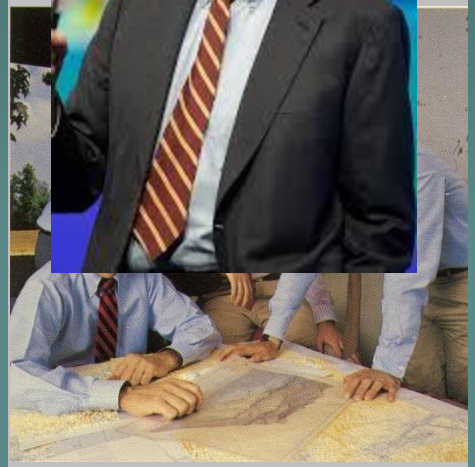


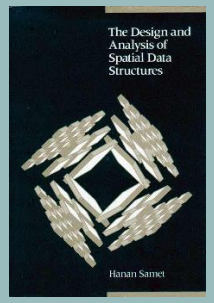
Kindly let me get the technology you have

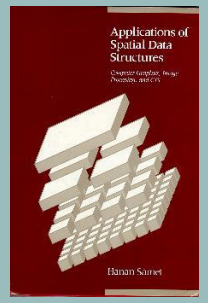




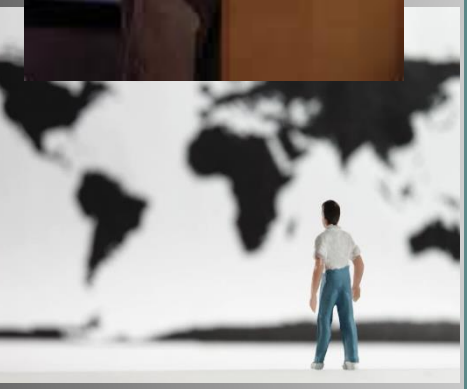


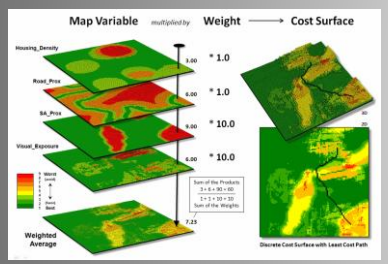


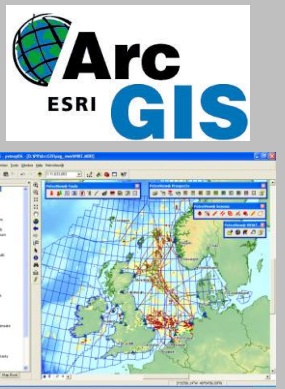

The Design and Analysis of Spatial Data Structures
Hanan Samet


Applications of Spatial Data Structures
Hanan Samet

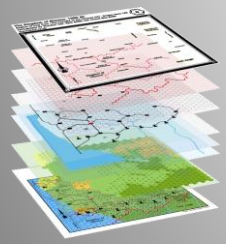

Spatial Databases
Manage GIS Data Efficiently

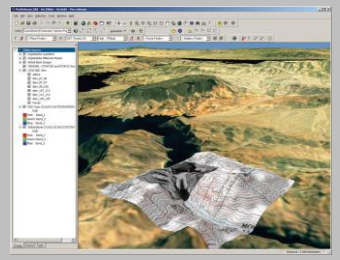


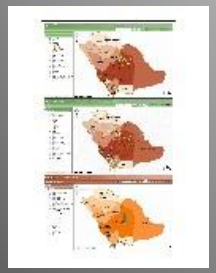

Map Variable
multiplied by
Weight
Cost Surface
Housing_Density
Road_Prox
BA_Prox
Visual_Exposure
* 1.0
* 1.0
* 10.0
* 10.0
Weighted Average
Discrete Cost Surface with Least Cost Path

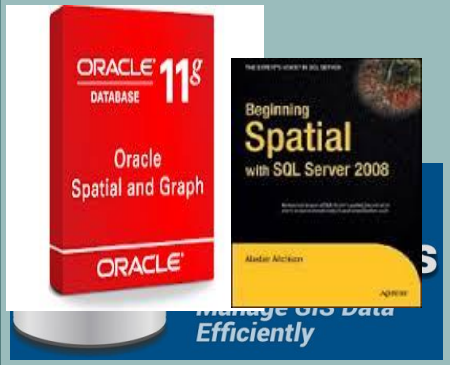


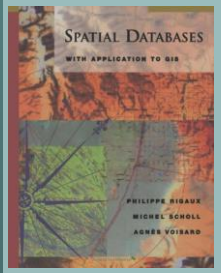

IRAN
Arabian Gulf
BAHRAIN
Manama
SAUDI ARABIA
QATAR
Doha
Riyadh
N
0 km 150

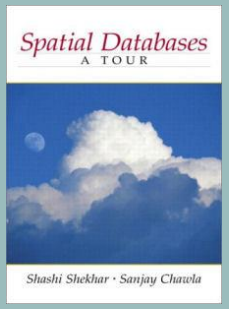


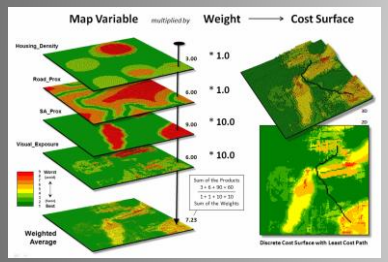




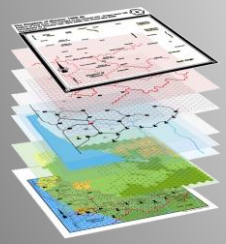

ORACLE 11g
DATABASE
Oracle
Spatial and Graph
ORACLE
Beginning
Spatial
with SQL Server 2008
Alastair Aitchison
Efficiently

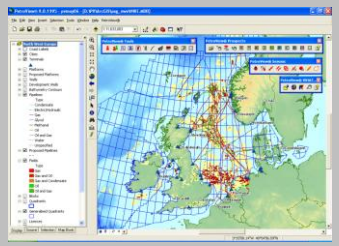

Spatial Databases
WITH APPLICATION TO GIS
PHILIPPE RIGAUX
MICHEL SCHOLL
AGNÈS VOISARD

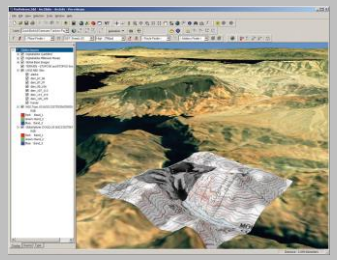

Spatial Databases
A TOUR
Shashi Shekhar · Sanjay Chawla

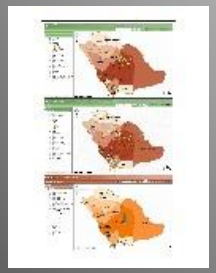

Map Variable
multiplied by
Weight
Cost Surface
Housing_Density
Road_Prox
BA_Prox
Visual_Exposure
* 1.0
* 1.0
* 10.0
* 10.0
Weighted Average
Discrete Cost Surface with Least Cost Path

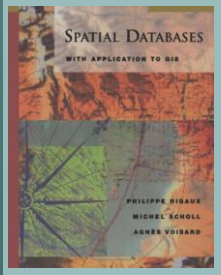

IRAN
Arabian Gulf
BAHRAIN
Manama
SAUDI ARABIA
QATAR
Doha
Riyadh
N
0 km 150

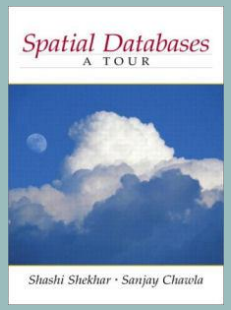




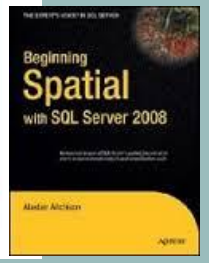





Spatial Databases
WITH APPLICATION TO GIS
PHILIPPE RIGAUX
MICHEL SCHOLL
AGNÈS VOISARD


Spatial Databases
A TOUR
Shashi Shekhar · Sanjay Chawla


ORACLE 11g
DATABASE
Oracle
Spatial and Graph
ORACLE


Beginning
Spatial
with SQL Server 2008
Alastair Aitchison


PostGIS
Geospatial Objects for PostgreSQL


Arc
ESRI
GIS


idrisi


ORACLE


IBM
DB2


Microsoft
SQL Server


PostgreSQL
the world's most advanced open source database


MySQL


SYBASE
An SAP Company

Google

bing

twitter

facebook

Map Reduce

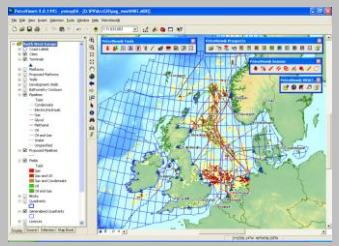
amazon web services™ EC2

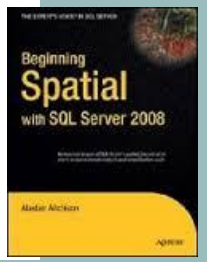
hadoop

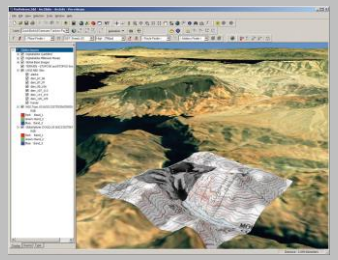
cloudera IMPALA Spark

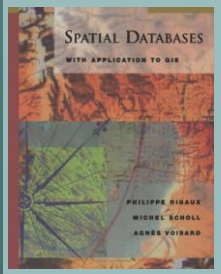
Arc GIS ESRI

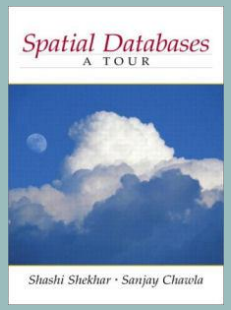
PostGIS
Geospatial Objects for PostgreSQL











ORACLE

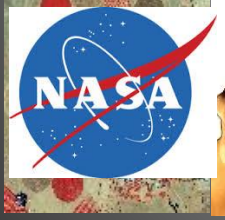
IBM DB2

Microsoft SQL Server

PostgreSQL

MySQL

SYBASE An SAP Company



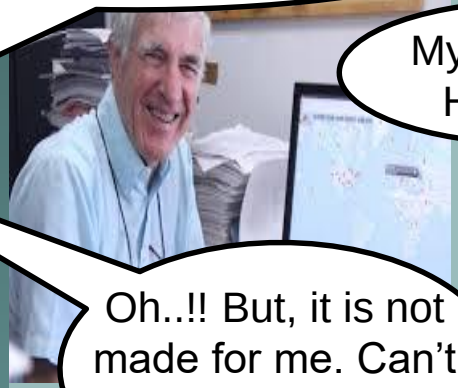
idrisi

Let me check with my other good friends there.

HELP..!! Again, I have **BIG** data. Your technology is not helping me

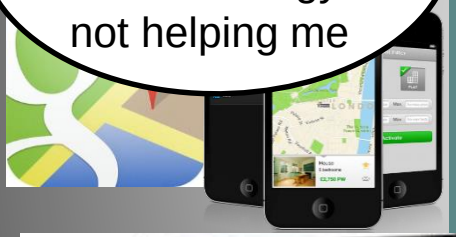
Cool Big Data technology..!! Can I use it in my application?

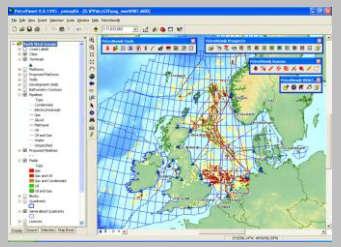
Sorry, seems like the DBMS technology cannot scale more

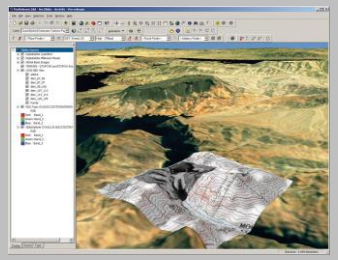


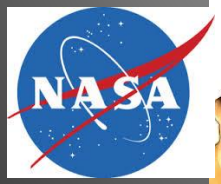
My pleasure. Here it is.

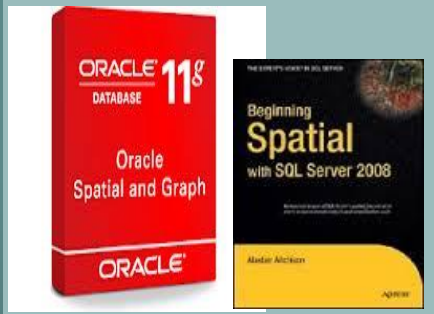
Oh..!! But, it is not made for me. Can't make use of it as is

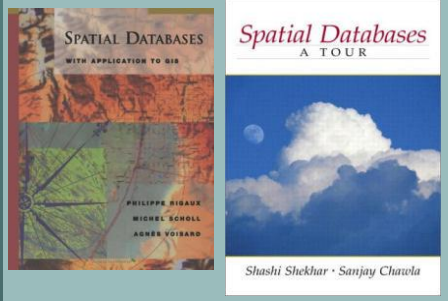












Google

bing

twitter

facebook

Map Reduce

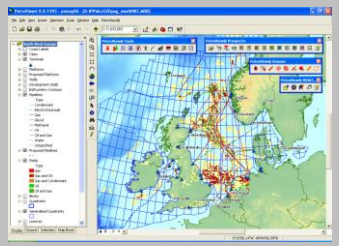
amazon web services™ EC2

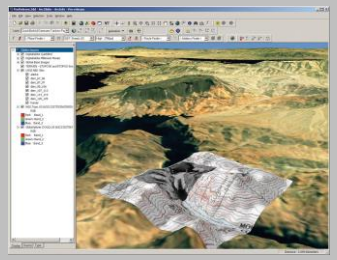
hadoop

cloudera IMPALA Spark

Arc GIS ESRI

PostGIS Geospatial Objects for PostgreSQL

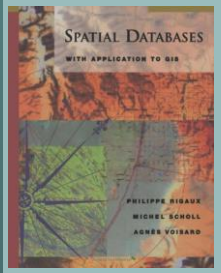


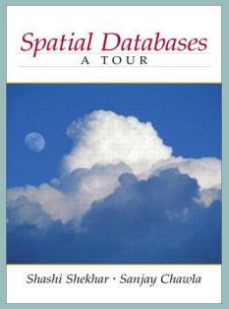


ORACLE 11g DATABASE Oracle Spatial and Graph ORACLE

Beginning Spatial with SQL Server 2008 Alastair Aitchison

idrisi





ORACLE

IBM DB2

Microsoft SQL Server

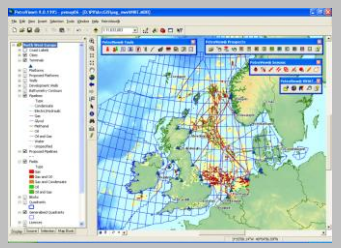
PostgreSQL

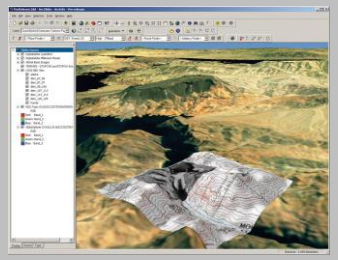
MySQL

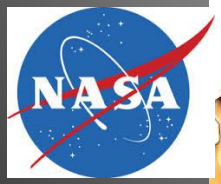
SYBASE An SAP Company

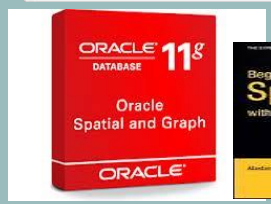

Kindly let me understand your needs
Kindly let me get the technology you have
ESRI ArcGIS
facebook
MapReduce
amazon web services
EC2
hadoop
cloudera IMPALA
Spark
NASA
idrisi
PostGIS
Geospatial Objects for PostgreSQL
ORACLE 11g DATABASE
Oracle Spatial and Graph
ORACLE
Beginning Spatial with SQL Server 2008
SPATIAL DATABASES WITH APPLICATION TO GIS
Spatial Databases A Tour
Shashi Shekhar · Sanjay Chawla
ORACLE
IBM DB2
Microsoft SQL Server
PostgreSQL
MySQL
SYBASE An SAP Company

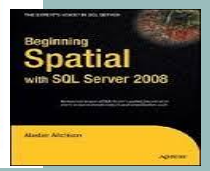
The Era of Big Spatial Data

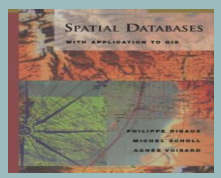


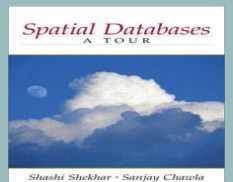












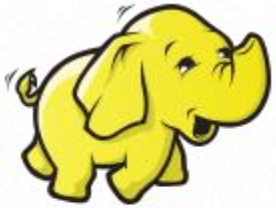
The Era of Big Spatial Data

UCR

Recent products are there….

GeoSpark

rasdaman
raster data manager



Hadoop-GIS
Spatial Big Data Solutions

geomesa

GeoTrellis

SpaceCurve

SPHINX

Spatial Computing

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies
  - [tentative] → emerging definition and field
  - Technologies could be software, hardware, or both

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies
  - [tentative] → emerging definition and field
  - Technologies could be software, hardware, or both
- Major questions of interest:
  - Where am I?

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies
  - [tentative] → emerging definition and field
  - Technologies could be software, hardware, or both
- Major questions of interest:
  - Where am I?
    - On Earth, in a mall, in a campus, in a plaza, inside a human body…etc

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies
  - [tentative] → emerging definition and field
  - Technologies could be software, hardware, or both
- Major questions of interest:
  - Where am I?
    - On Earth, in a mall, in a campus, in a plaza, inside a human body…etc
  - What is around me?
    - restaurants, hotels, gas stations, ATMs…etc

Spatial Computing

- A field that innovates a set of technologies and techniques to combine spatial information with computing technologies
  - [tentative] → emerging definition and field
  - Technologies could be software, hardware, or both
- Major questions of interest:
  - Where am I?
    - On Earth, in a mall, in a campus, in a plaza, inside a human body…etc
  - What is around me?
    - restaurants, hotels, gas stations, ATMs…etc
  - What is in or around certain area(s)? (Spatial Analysis)
    - Situation after a natural disaster, changes over time, etc
    - Science, e.g., vegetation analysis, environment, ecology,…etc
    - Enterprise, e.g., agriculture, ride sharing, market research,…etc

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
  - Estimated value by 2020: $600B
    (McKinsey Global Institute, 2011 report on Big Data)

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
  - Estimated value by 2020: $600B
    (McKinsey Global Institute, 2011 report on Big Data)

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
- The governments

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
- The governments

May 18th, 2011

Congressional Research Service

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

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

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

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
- The governments
- The public

Who use Spatial Computing?

- Hundreds of millions of people (if not billions)
- Business
- The governments
- The public

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Applications for Emissions and Vehicles

- Two example applications:
  1. Discovering co-occurrence patterns in non-compliant emissions time intervals
  2. Energy-efficient vehicle path selection

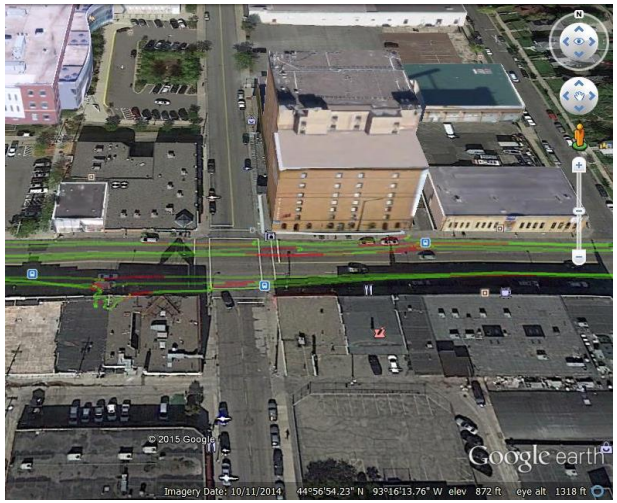
Discovering co-occurrences in non-compliant emissions time intervals

- **Eco-friendly transportation**
  - Despite stricter regulatory standards, vehicles are emitting at rates higher than their certified limit [6,7]
    - Tests do not accurately reflect the non-compliant real-world vehicle behavior.
- 

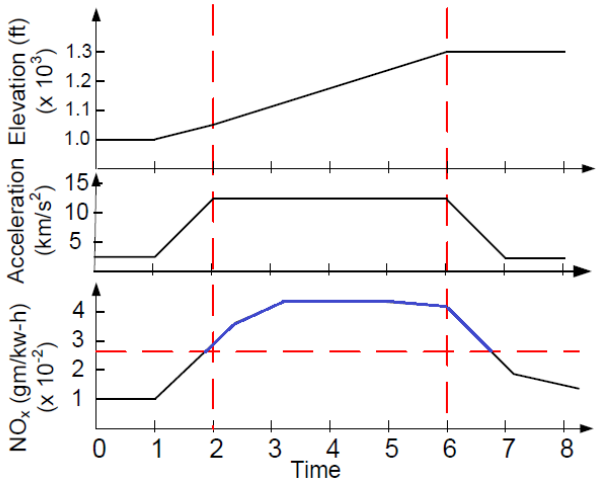
U.S. Fines Hyundai, Kia for Fuel Claims

Discovering co-occurrences in non-compliant emissions time intervals

- **Availability of Spatio-Temporal Big data**
  - e.g. GPS traces, engine measurements
  - ***USDOT RDT Strategic Plan 2013-18***[5]: "*Utilize data from vehicles on emissions and fuel consumption to better manage the transportation network to enable reduction in GHGs*"
  - What real-world "sub-journeys" experience non-compliant emissions or low fuel efficiency?



Non-compliant NO_x emissions in red



Candidate Co-occurrence patterns

- **Other apps:** e.g. Industrial process control, climate change

Problem Overview

- **An event:** e.g. e_1: wheelspeed ϵ [0, 5) km/h, e_2: wheelspeed ϵ [5, 10) km/h
  - variable v falls within range [vi,vi+1).

- **A multivariate event trajectory (MET):**

Explanatory variables

Target variable

Time	0	1	2	3	4	5	6	7
v_1: Engine Power	a_1	a_2	a_3	a_2	a_1	a_2	a_3	a_2
v_2: Engine RPM	b_1	b_1	b_2	b_3	b_1	b_1	b_2	b_2
NO_x (gm/sec)	0.21	0.3	0.35	0.28	0.2	0.15	0.1	0.1

  - a sequence of multi-variate points
  - Points defined over explanatory variables and a target variable

- **An event-sequence S(v):** e.g. $a_2a_3a_2$
  - a sequence of temporally contiguous events

Problem Overview

- **A non-compliant window (W_N)**:
  - a temporal window where target variable exceeds a given standard

e.g. Windows of length 3 where average NOx > 0.3

Time	0	1	2	3	4	5	6	7
v_1: Engine Power	a_1	a_2	a_3	a_2	a_1	a_2	a_3	a_2
v_2: Engine RPM	b_1	b_1	b_2	b_3	b_1	b_1	b_2	b_2
NO_x (gm/sec)	0.21	0.3	0.35	0.28	0.2	0.15	0.1	0.1

Which windows are non-compliant?

Problem Overview

- **A Non-compliance Sub-time-series Co-occurrence Pattern (NSC):**
  - A set of event-sequences **within a time lag δ** from a non-compliant window
  - Defined only on **explanatory** variables
  - Sequences are **equal** in length

e.g.

δ =1 sec

Time	0	1	2	3	4	5	6	7
v_1: Engine Power	a_1	a_2	a_3	a_2	a_1	a_2	a_3	a_2
v_2: Engine RPM	b_1	b_1	b_2	b_3	b_1	b_1	b_2	b_2
NO_x (gm/sec)	0.21	0.3	0.35	0.28	0.2	0.15	0.1	0.1

ID	Candidate NSC Patterns	Dim
1	$\{a_2a_3a_2\}$	1
2	$\{b_1b_2b_3\}$	1
3	$\{a_2a_3a_2, b_1b_2b_3\}$	2
4	$\{a_1a_2a_3\}$	1
5	$\{b_1b_1b_2\}$	1
6	$\{a_1a_2a_3, b_1b_1b_2\}$	2

Problem Overview

Interest Measure: Cross-K function

- how much the association between pattern C and non-compliant windows W_N at lag δ deviates from independence?

$$\hat{K}_{C,W_N}(\delta) = \frac{T \times |C \overset{\delta}{\bowtie} W_N|}{|W_N||C|}, \quad T = \sum_{all\ METS} \tau$$

$|W_N|$: number of non-compliant windows

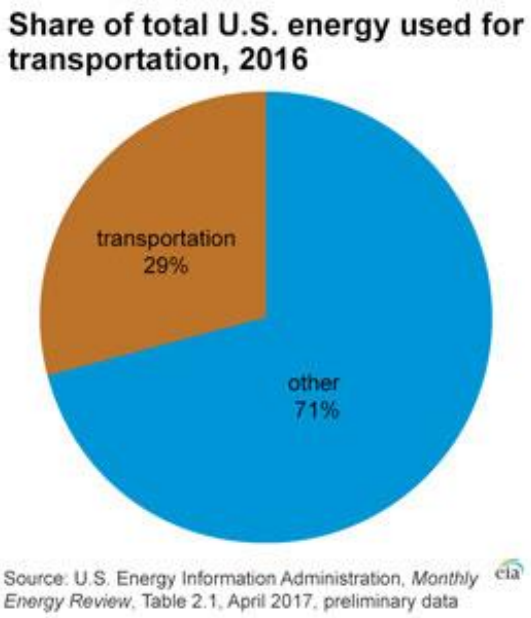
$|C|$: pattern cardinality

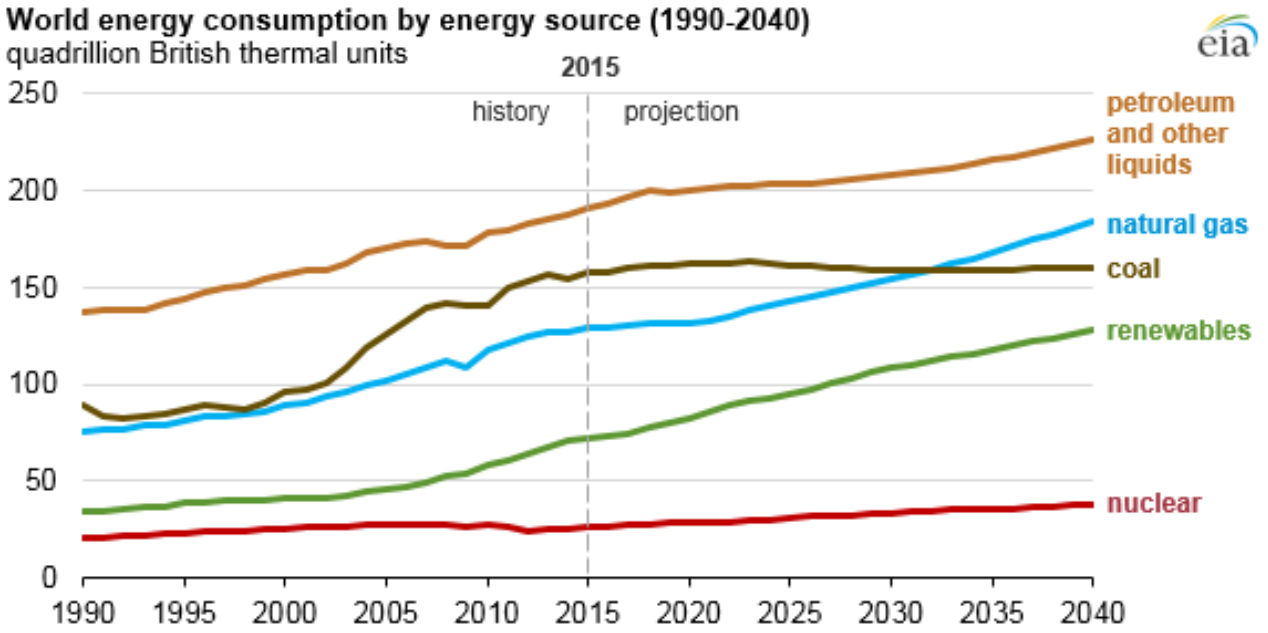
$|C \bowtie W_N|$: Join set cardinality between instances of C and W_N at δ

Under independence:
$\hat{K}_{C,WN}(\delta) = \delta + 1$

Energy-efficient Vehicle Path Selection

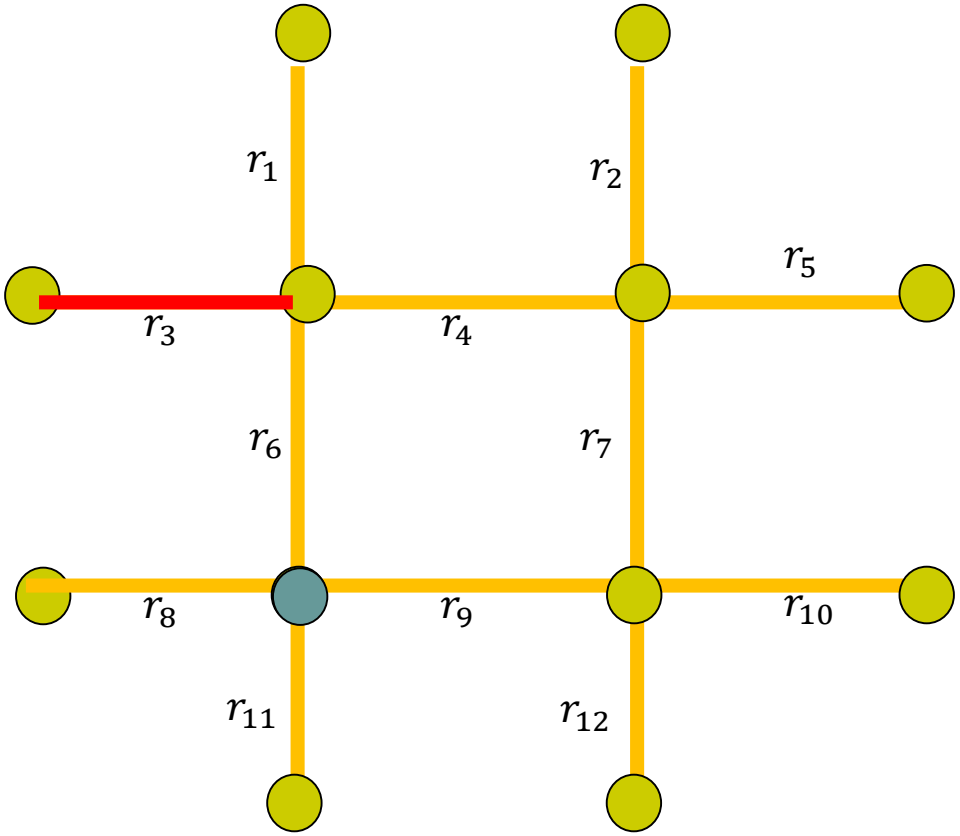
- In 2015, more than $1.237 trillion was spent on energy in the U.S., of which the major user is transportation.
- The energy consumption will rise 28% between 2015 and 2040, among which the share of transportation is about 29%.





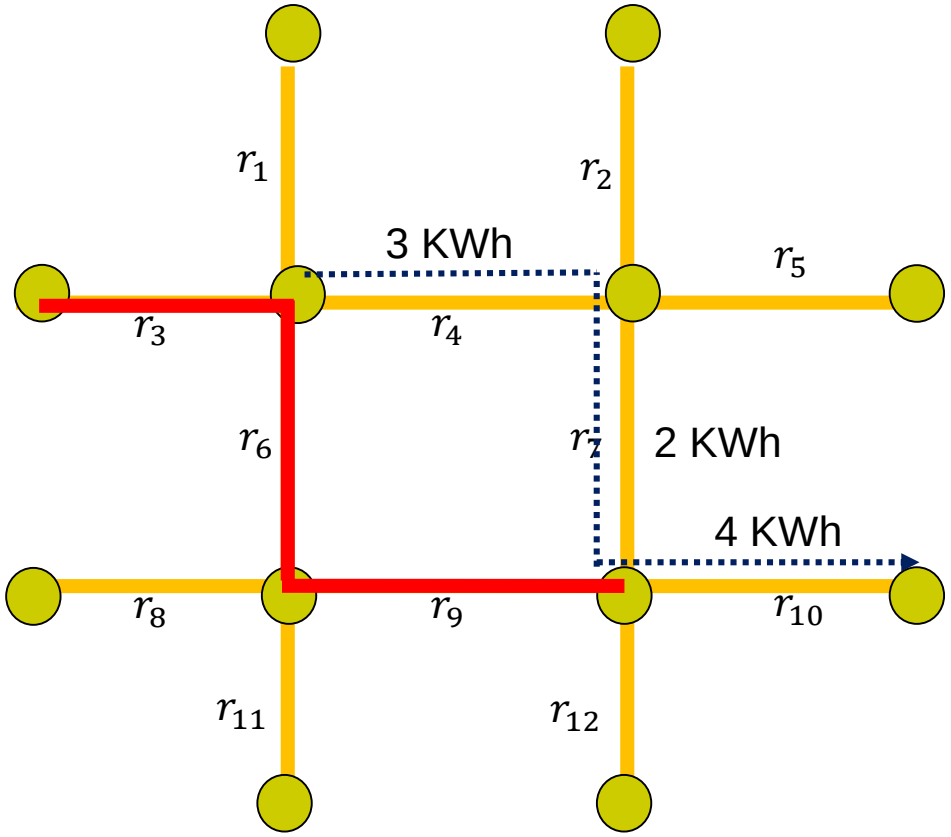
Problem Overview

- Road network
  - Segment
  - Intersection



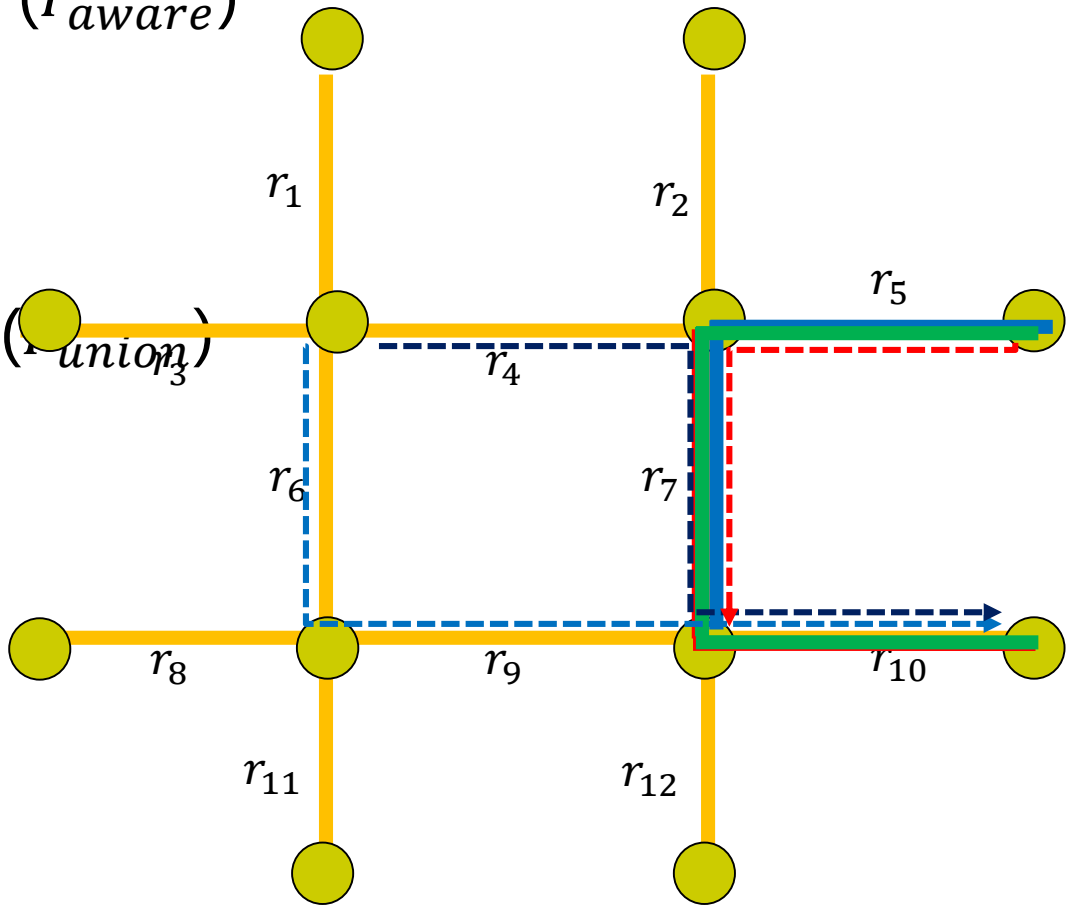
Problem Overview

- Path
- Trajectory
  - Path: $[r_4, r_7, r_{10}]$
  - Energy consumption: [3,2,4] (KWh)



Problem Overview

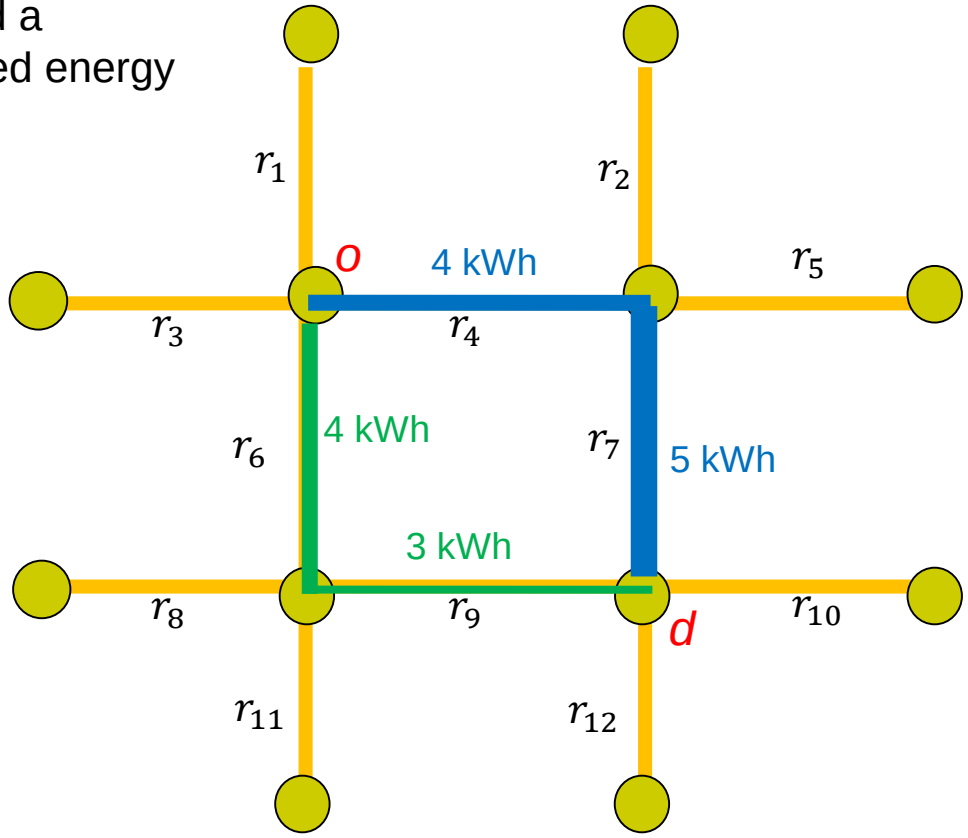
- Trajectory-aware path (P_{aware})
  - $[r_5, r_7]$
  - $[r_7, r_{10}]$
- Trajectory-union path (P_{union})
  - $[r_5, r_7, r_{10}]$



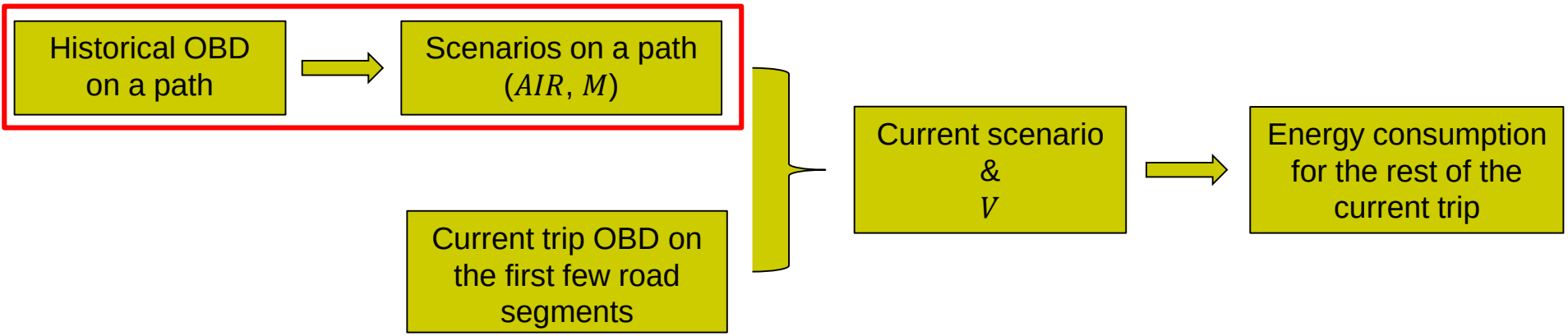
B. Yang, J. Dai, C. Guo, C. S. Jensen, and J. Hu, “PACE: a PAth-CEntric paradigm for stochastic path finding,” *The VLDB Journal*, vol. 27, no. 2, pp. 153–178, Apr. 2018.

Problem Overview

- Energy-efficient path:
  - A path between an origin and a destination with least expected energy consumption.



Energy Consumption Estimation



Assumption: the current trip is short, and the factors affecting it velocity pattern do not change.

$$W = \int \left(\frac{1}{2\eta} c_{air} A \rho v^3\right) dt + \frac{m}{\eta} \int (av + c_{rr} g v) dt$$

energy for air resistance (AIR) — vehicle parameter (V) — motion property (M)

$$W = AIR + V \times M$$

Onboard diagnostic data (OBD):

- **355** engine measurement fields
  - Timestamp
  - Spatial information (Longitude, Latitude, Altitude)
  - Vehicle information (e.g., Motor Speed, Battery Current, Energy used)

Challenges in Geospatial Tech

Challenges: Privacy vs. Utility

- Check-in risks: Stalking, GeoSlavery, Others know that you are not home, etc

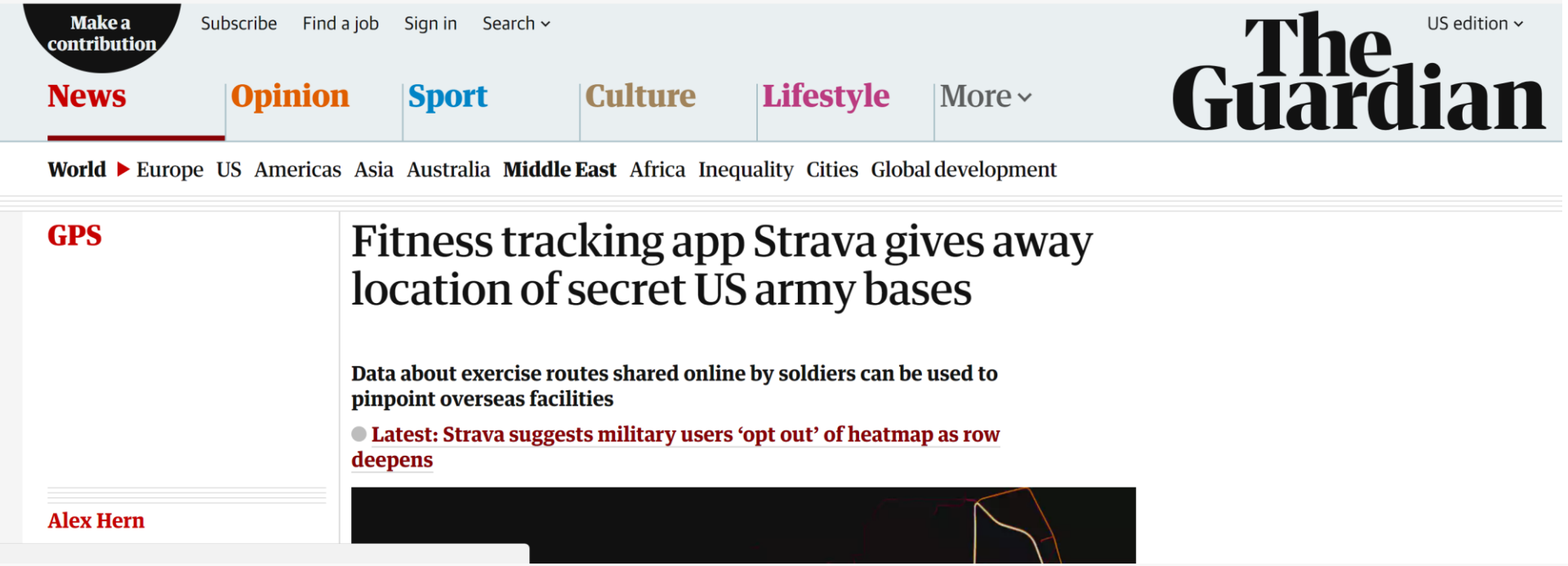
Challenges: Privacy vs. Utility

- Check-in risks: Stalking, GeoSlavery, Others know that you are not home, etc
- Ex: Girls Around me App (3/2012)

The Girls of Girls Around Me. It's doubtful any of these girls even know they are being tracked. Their names and locations have been obscured for privacy reasons. (Source: Cult of Mac, March 30, 2012)

Challenges: Security vs. Utility

Challenges: Security vs. Utility



Make a contribution Subscribe Find a job Sign in Search US edition

The Guardian

News Opinion Sport Culture Lifestyle More

World Europe US Americas Asia Australia Middle East Africa Inequality Cities Global development

GPS

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users 'opt out' of heatmap as row deepens

Alex Hern

Location-based threats: How cybercriminals target you based on where you live

Corporate · Network · Security Tips · SophosLabs · Cryptowall · Geomalware · Locky · Phishing · Ransomware · Sophos Home · Spam · TorrentLocker

Challenges: Security vs. Utility

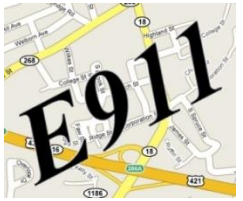
- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?

Challenges: Security vs. Utility

- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?
- Involved groups:
  - Civil Society
  - Economic Entities
  - Public Safety
  - Policy Makers

Challenges: Security vs. Utility

- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?
- Involved groups:
  - Civil Society
  - Economic Entities
  - Public Safety
  - Policy Makers
- Agreements and disagreements
  - Agreements: E911, emergency alerts
  - Controversial: traffic monitoring



Spatial beyond GeoSpatial

- Examples:
  - Human bodies
  - VLSI chips and boards
  - Universe
  - Indoor and virtual spaces

Spatial beyond GeoSpatial

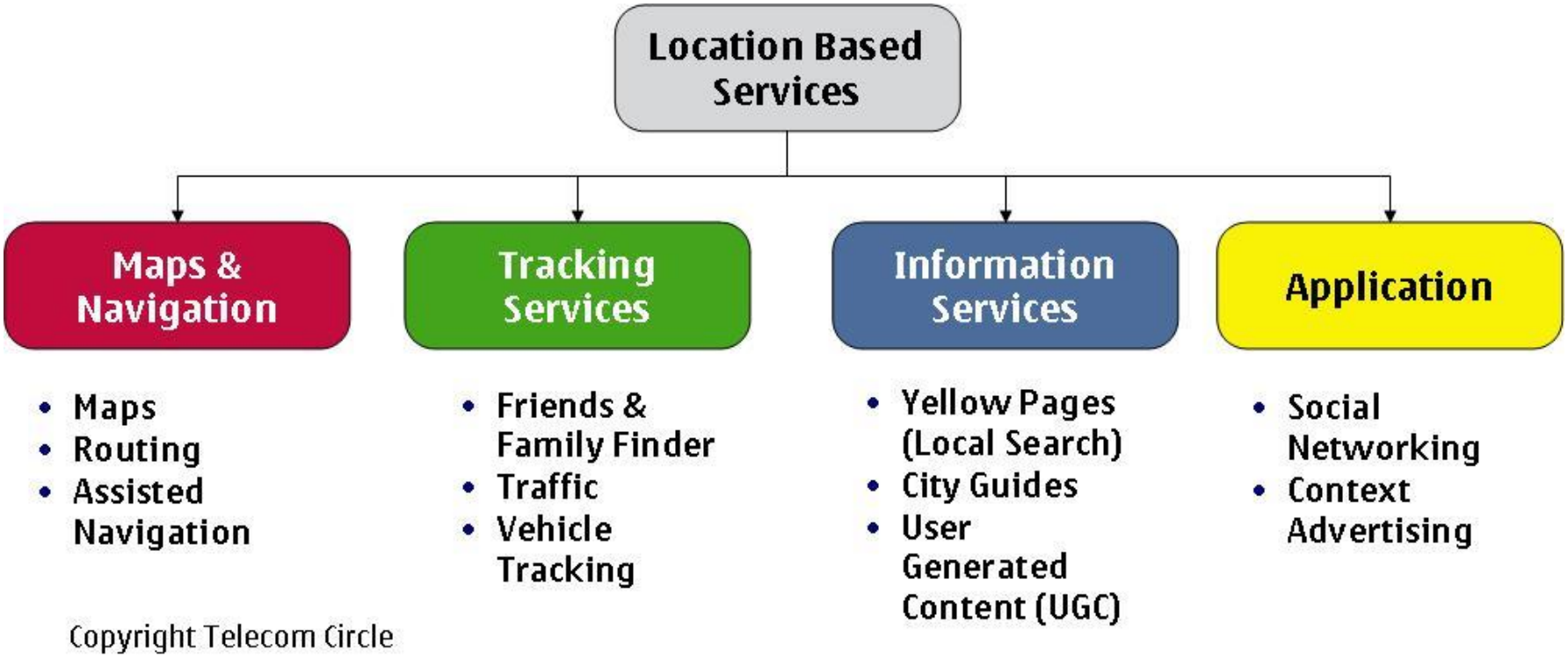
- Examples:
  - Human bodies
  - VLSI chips and boards
  - Universe
  - Indoor and virtual spaces
- Challenges:
  - What are the reference system?
    - On Mars? Outside Milkyway galaxy? In augmented reality spaces?
    - Is it one for all humans? Or personalized?
  - Accuracy
  - 3D+ scalability

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Location Based Services

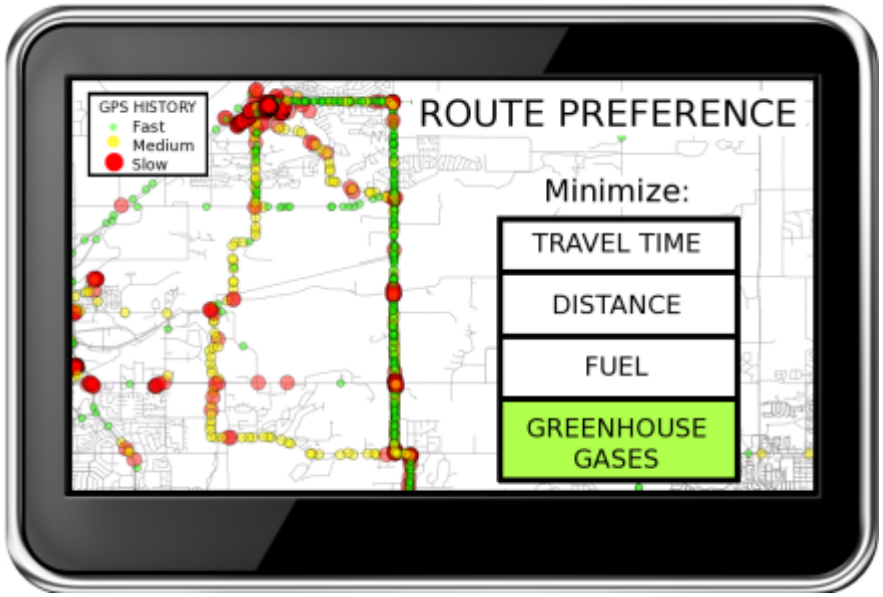
- Services based on your location
  - Location Sharing: Where am I? (street address, <latitude, longitude>)
  - Directory: Where is the nearest gas station?
  - Routes: What is the shortest path to reach there?



Copyright Telecom Circle

Trends: Next Generation Navigation

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



Trends: Persistent Geo-Hazard Monitoring

- Environmental influences on our health & safety
  - air we breathe, water we drink, food we eat

Trends: Persistent Geo-Hazard Monitoring

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

References and Credits

- References
  - CACM Article: https://cacm.acm.org/magazines/2016/1/195727-spatial-computing/fulltext
  - CCC Workshop Report: https://cra.org/wp-content/uploads/sites/2/2015/05/Spatial_Computing_Report-2013.pdf
  - Spatial Computing Lectures: https://www.youtube.com/watch?v=ftwWfB7JWaQ&list=PLq_27Uv53bDm3hyXd5QWG-N8L4Vgvcy9J&index=1
- Credits:
  - Prof. Ahmed Eldawy and Prof. Mohamed Mokbel tutorial
    - http://www.vldb.org/pvldb/vol10/p1992-eldawy.pdf
  - Prof. Shashi Shekhar book slides
    - http://www.spatial.cs.umn.edu/Book/slides/
  - Reem Ali paper slides (Discovering Sub-time-series Co-occurrence Patterns of Non-compliance)
  - Yan Li paper slides (Physics-guided Energy-efficient Path Selection)

UNIVERSITY OF CALIFORNIA
UC RIVERSIDE

Thank You

Questions?

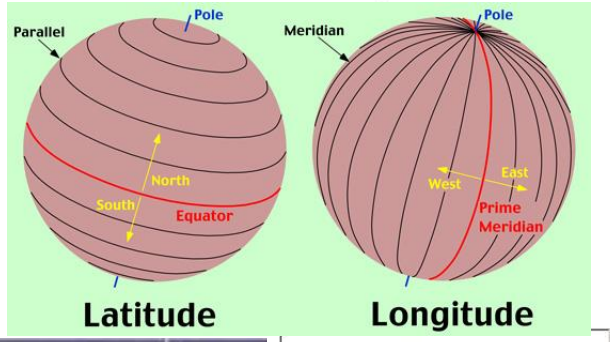
Email: amr@cs.ucr.edu
www.cs.ucr.edu/~amr/

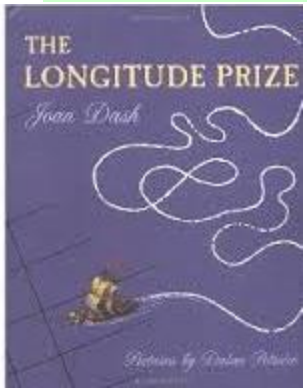
Major technologies and areas (past, present, & future)

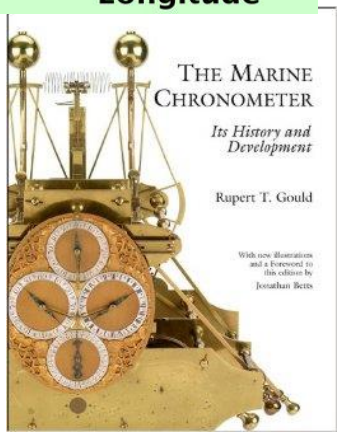
- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Global Positioning Systems (GPS)

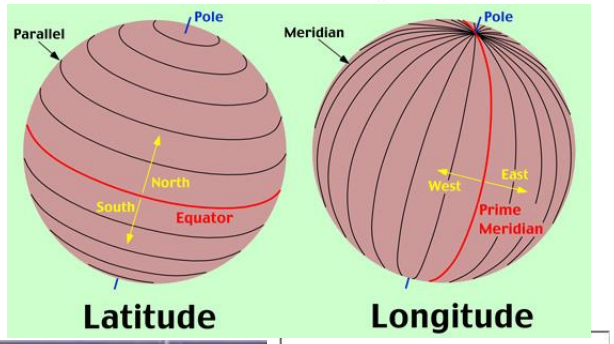
- Positioning ships
  - Latitude f(compass, star positions) → ancient and medieval civilizations
  - Longitude Prize (1714) → marine chronometer



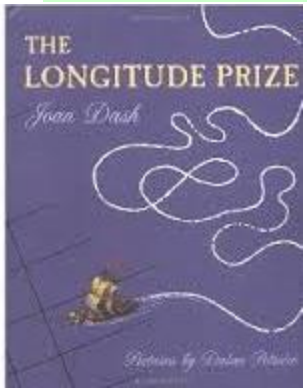


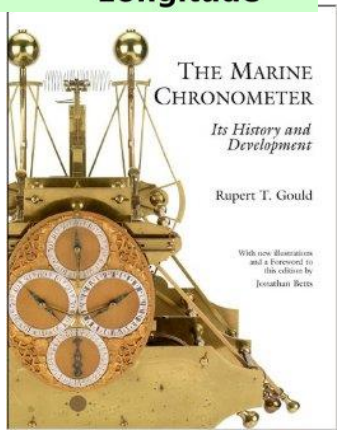


Global Positioning Systems (GPS)

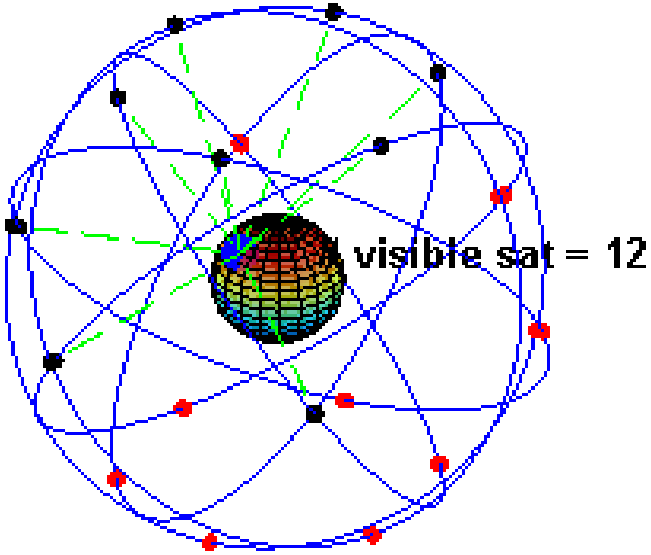


- Positioning ships
  - Latitude f(compass, star positions) → ancient and medieval civilizations
  - Longitude Prize (1714) → marine chronometer



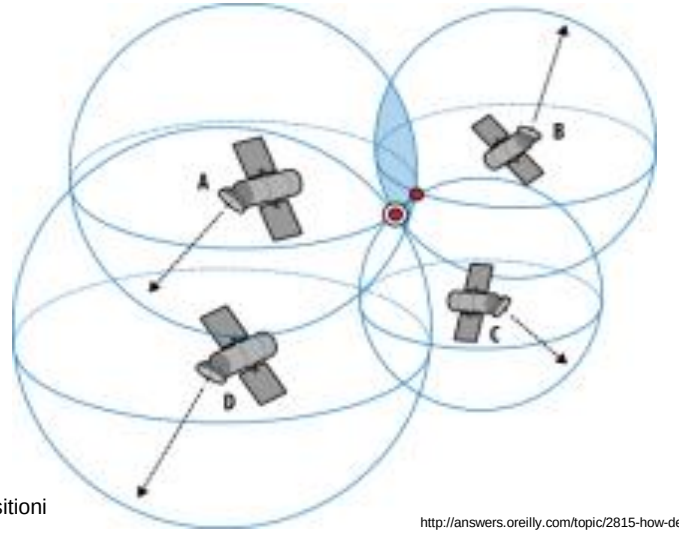


- Global Navigation Satellite Systems
  - Infrastructure: satellites, ground stations, receivers, …
  - Use: Positioning (sub-centimeter), Clock synchronization

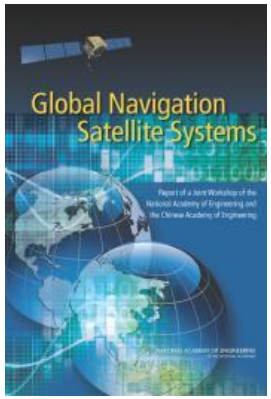


http://en.wikipedia.org/wiki/Global_Positioning_System

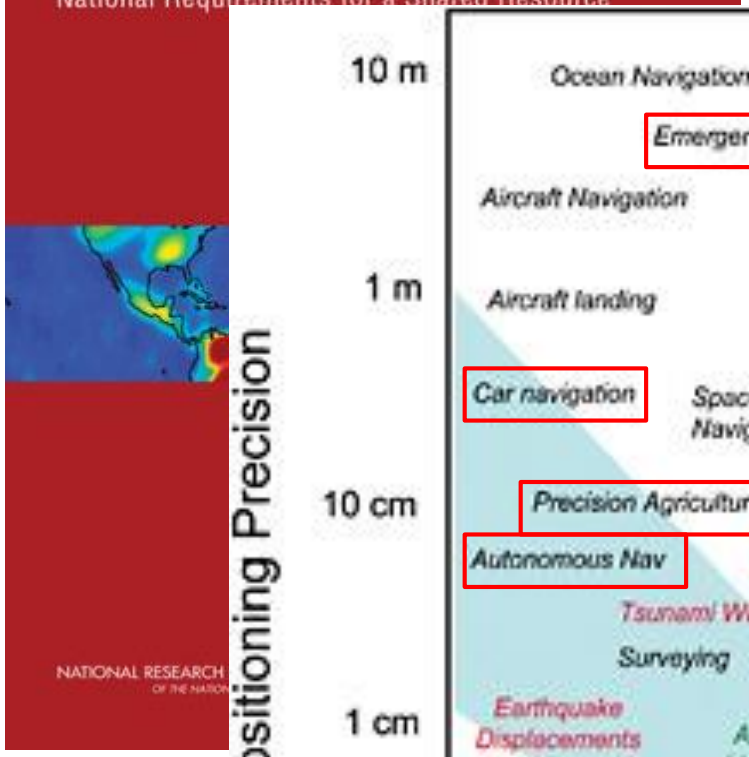
Trilateration



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



Positioning Precision



Future & Trends: Localization Indoors, Underground, & Underwater

- GPS works outdoors, but,
  - We are indoors 90% of time!
  - Ex. malls, hospitals, airports, …

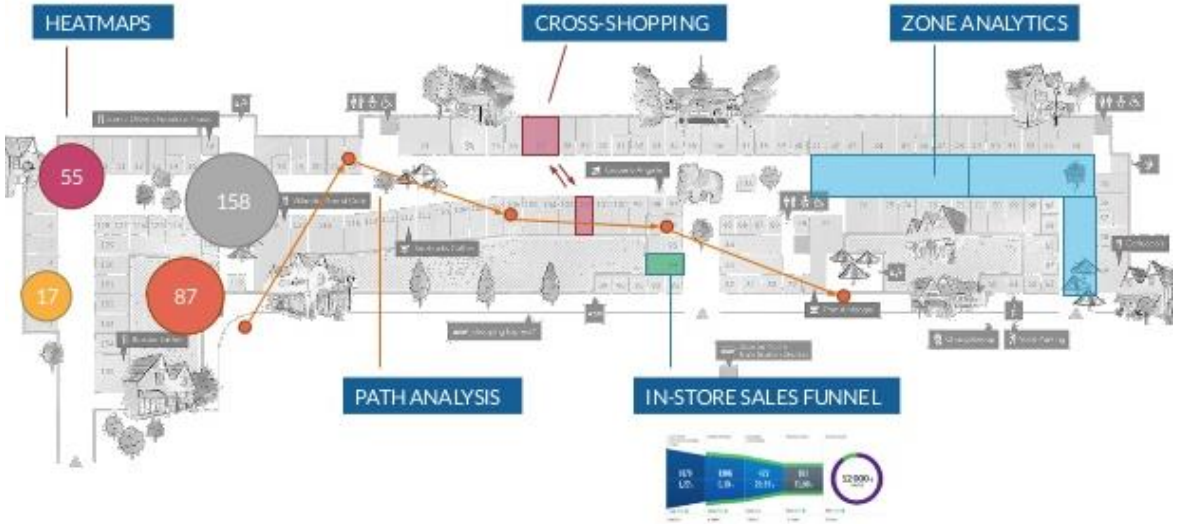
Future & Trends: Localization Indoors, Underground, & Underwater

- GPS works outdoors, but,
  - We are indoors 90% of time!
  - Ex. malls, hospitals, airports, …

Future & Trends: Localization Indoors, Underground, & Underwater

- GPS works outdoors, but,
  - We are indoors 90% of time!
  - Ex. malls, hospitals, airports, …
- Leveraging existing indoor infrastructure
  - Blue Tooth, Wi-Fi, …

Get In-Store Notifications

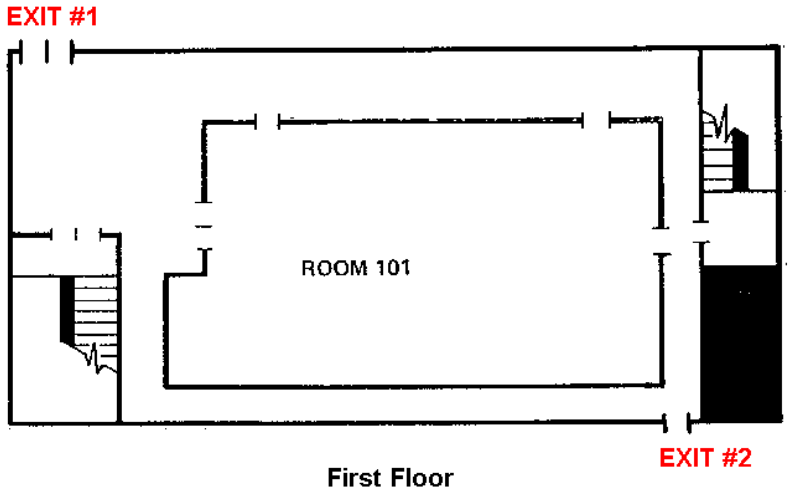


Future & Trends: Localization Indoors, Underground, & Underwater

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

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

Get In-Store Notifications



WiFi Localization

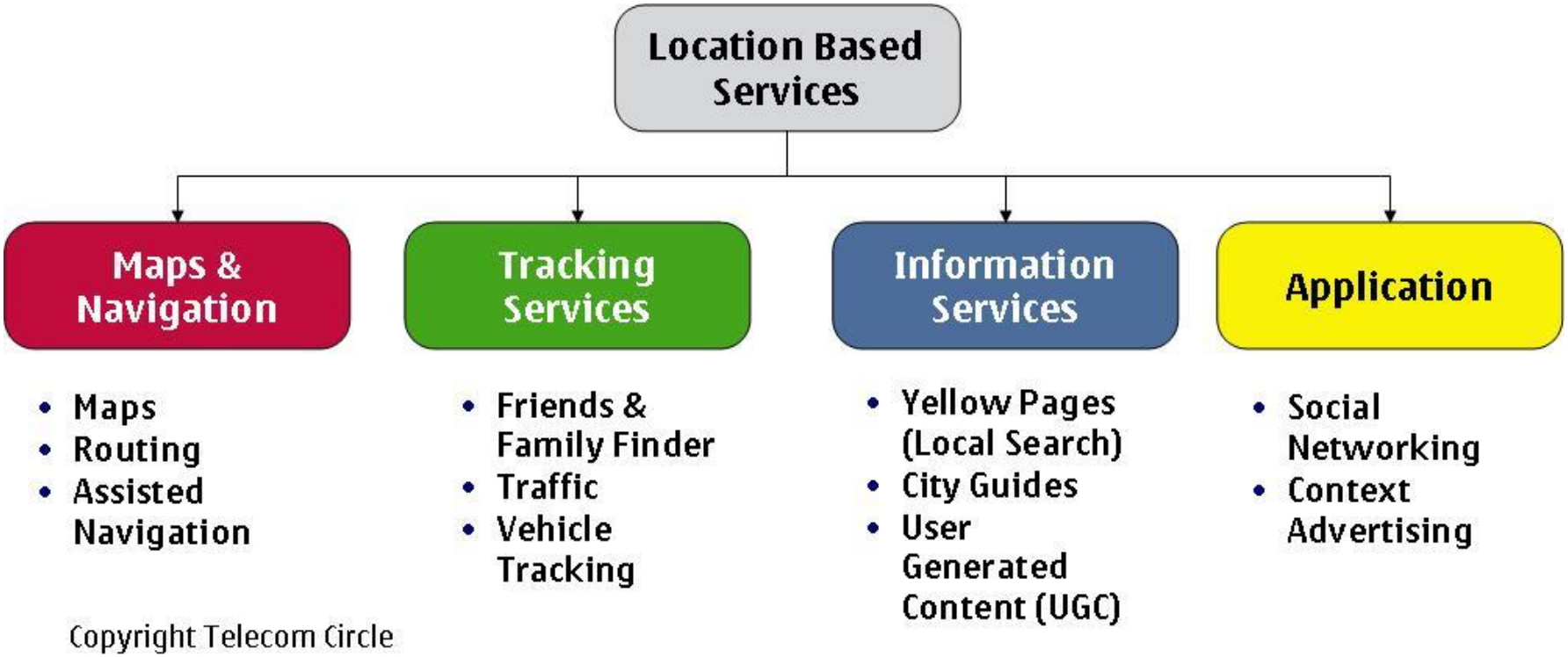
http://rfid.net/basics/rtls/123-wi-fi-how-it-works

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Location Based Services

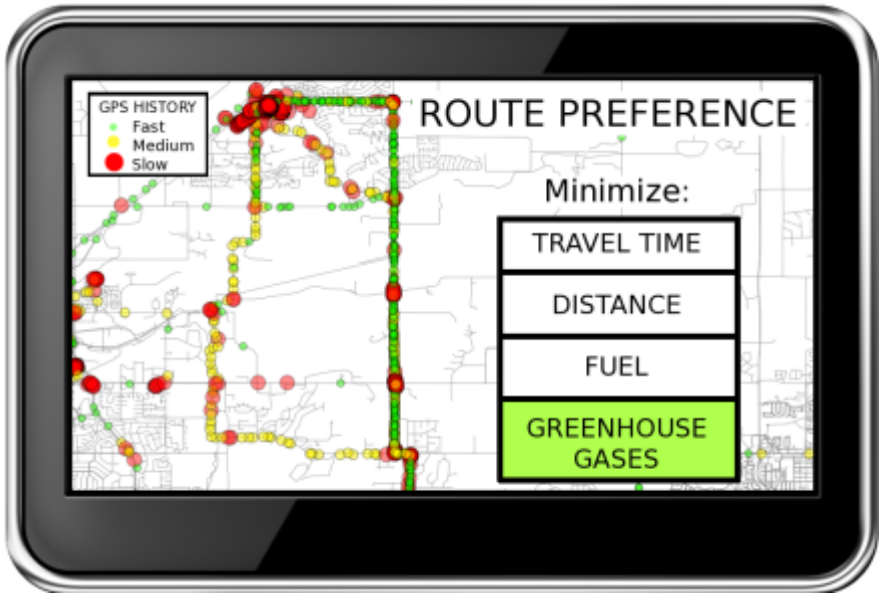
- Services based on your location
  - Location Sharing: Where am I? (street address, <latitude, longitude>)
  - Directory: Where is the nearest gas station?
  - Routes: What is the shortest path to reach there?



Copyright Telecom Circle

Trends: Next Generation Navigation

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



Trends: Persistent Geo-Hazard Monitoring

- Environmental influences on our health & safety
  - air we breathe, water we drink, food we eat

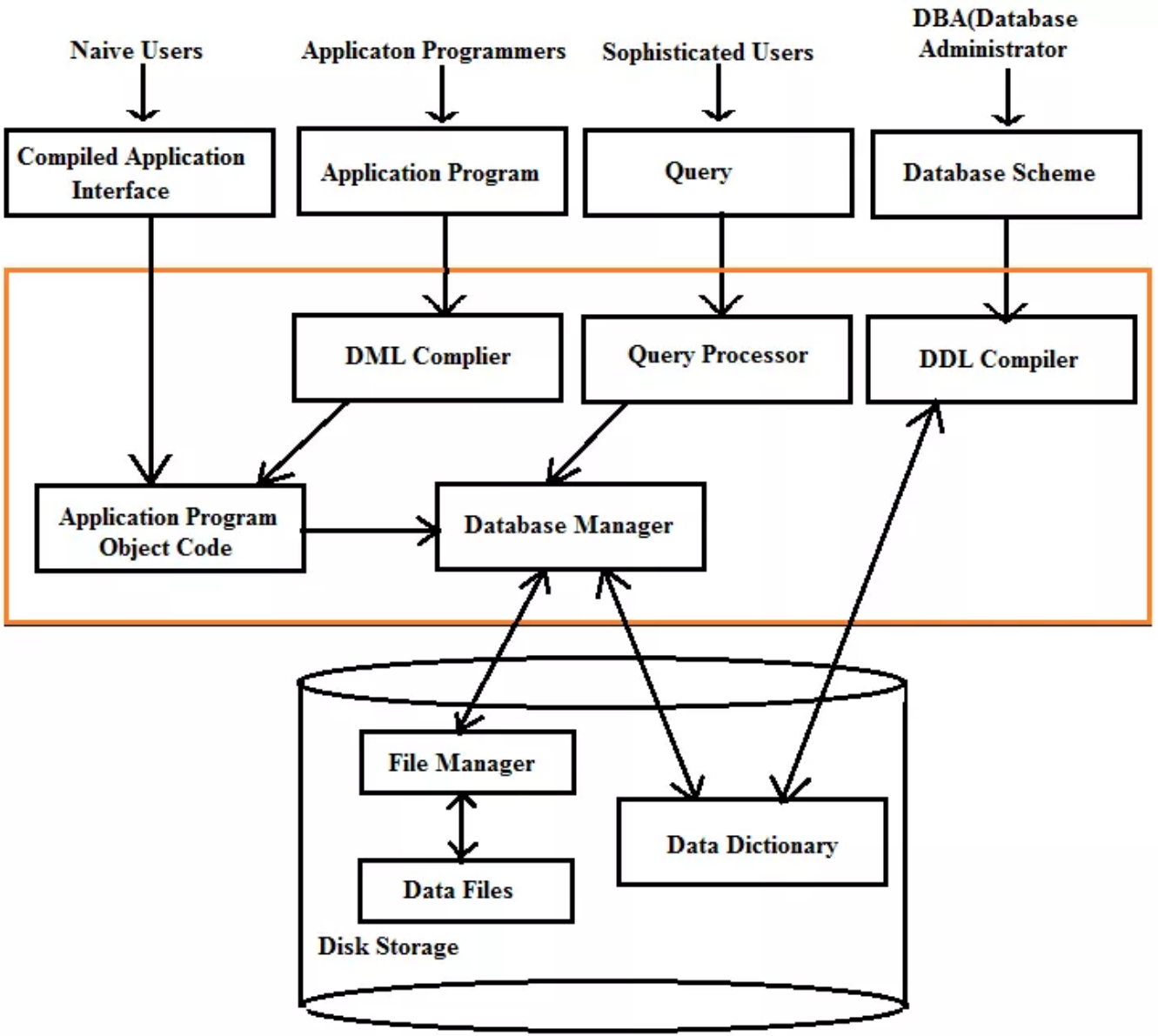
Trends: Persistent Geo-Hazard Monitoring

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

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Database Management Systems (DBMSs)



Spatial Database Management Systems (SDBMS)

- An SDBMS is a software module that:
  - Can work with an underlying database management system (DBMS)
  - Supports spatial data models, spatial abstract data types (ADTs) and a query language from which these ADTs are callable

Spatial Database Management Systems (SDBMS)

- An SDBMS is a software module that:
  - Can work with an underlying database management system (DBMS)
  - Supports spatial data models, spatial abstract data types (ADTs) and a query language from which these ADTs are callable
  - Supports spatial indexing, efficient algorithms for processing spatial operations, and domain specific rules for query optimization

SDBMS: Spatial Data Examples

- Examples of non-spatial data
  - Names, phone numbers, email addresses of people

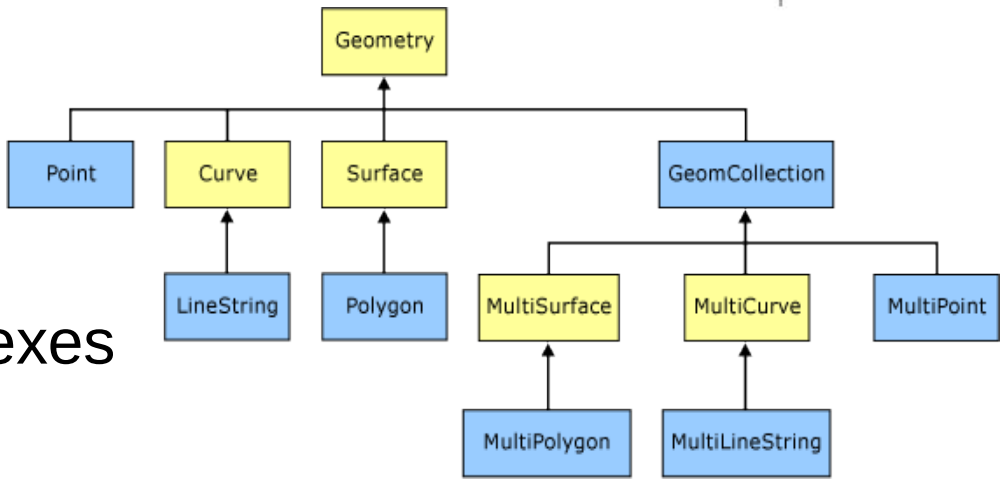
- Examples of spatial data
  - Census Data
  - NASA satellites imagery - terabytes of data per day
  - Weather and climate data
  - Rivers, farms, ecological impact
  - Medical imaging

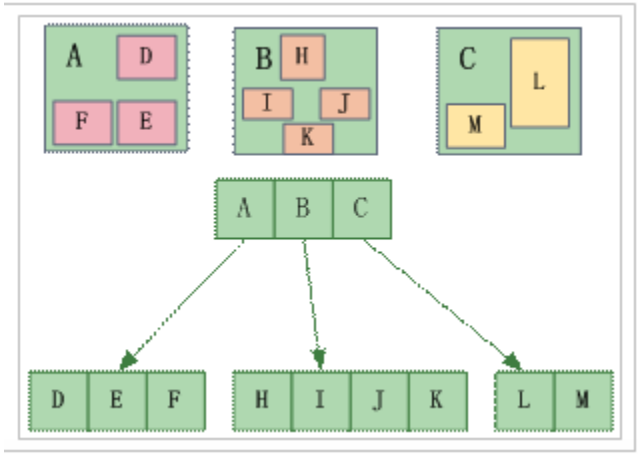
SDBMS: Non-Spatial vs. Spatial Queries

- Non-spatial queries
  - List the names of all bookstore with more than ten thousand titles
  - List the names of ten customers, in terms of sales, in the year 2001
- Spatial Queries
  - List the names of all bookstores with ten miles of Minneapolis
  - List all customers who live in Tennessee and its adjoining states

Components of an SDBMS

- Spatial data model
- Query language
- Query processing
- File organization and indexes
- Query optimization, etc.





SDBMS Example

- Consider a spatial dataset with:
  - County boundary (dashed white line)
  - Census block - name, area, population, boundary (dark line)
  - Water bodies (dark polygons)
  - Satellite Imagery (gray scale pixels)

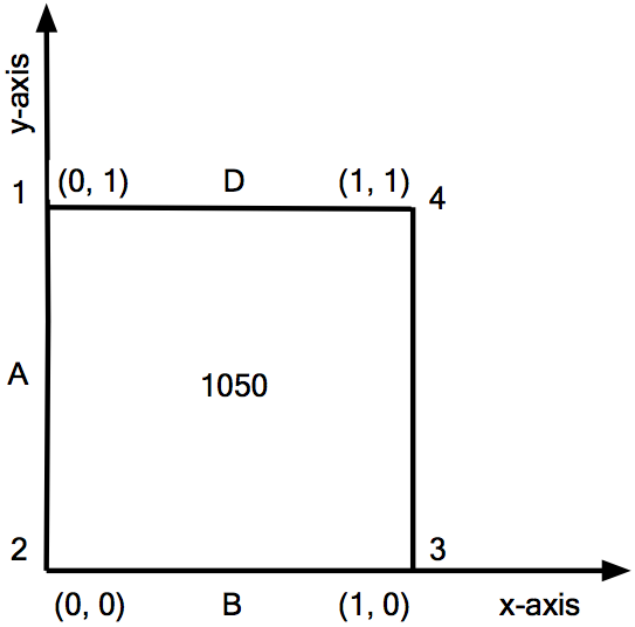
SDBMS Example

- Consider a spatial dataset with:
  - County boundary (dashed white line)
  - Census block - name, area, population, boundary (dark line)
  - Water bodies (dark polygons)
  - Satellite Imagery (gray scale pixels)
- Storage in a SDBMS table:

```
create table census_blocks (
name            string,
area            float,
population      number,
boundary        polygon );
```

SDBMS Example

- A row in the table **census_blocks**
- Boundary has a spatial data type that can be manipulated by the query language, query processor, indexes, etc

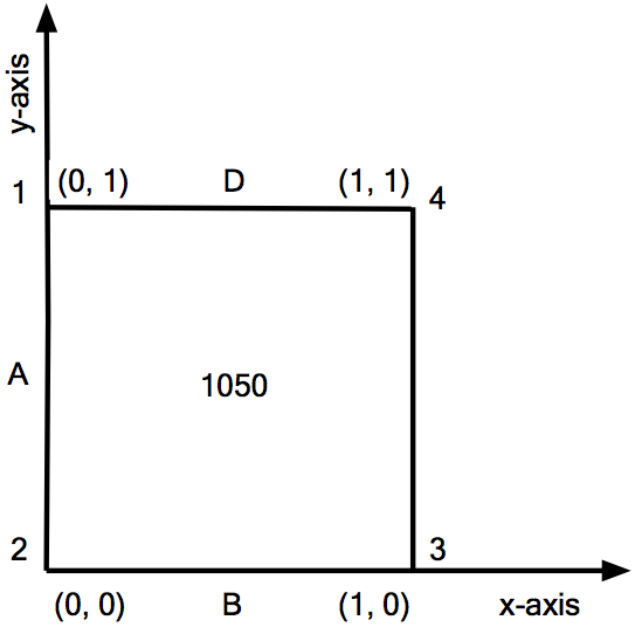


Census_blocks

Name	Area	Population	Boundary
1050	1	1839	Polyline ((0,0), (0,1), (1,1), (1,0))

SDBMS Example

- A row in the table **census_blocks**
- Boundary has a spatial data type that can be manipulated by the query language, query processor, indexes, etc
- Query: Select * FROM census_blocks C, factory F WHERE Overlap(C.boundary, F. boundary)



Census_blocks

Name	Area	Population	Boundary
1050	1	1839	Polyline ((0,0), (0,1), (1,1), (1,0))

Spatial beyond Databases

- Distributed systems
  - Hadoop, Spark, Impala, …etc

Spatial beyond Databases

GeoSpark

rasdaman
raster data manager

Hadoop-GIS
Spatial Big Data Solutions

GeoTrellis

SpaceCurve

SPHINX

Challenges: Privacy vs. Utility

- Check-in risks: Stalking, GeoSlavery, Others know that you are not home, etc

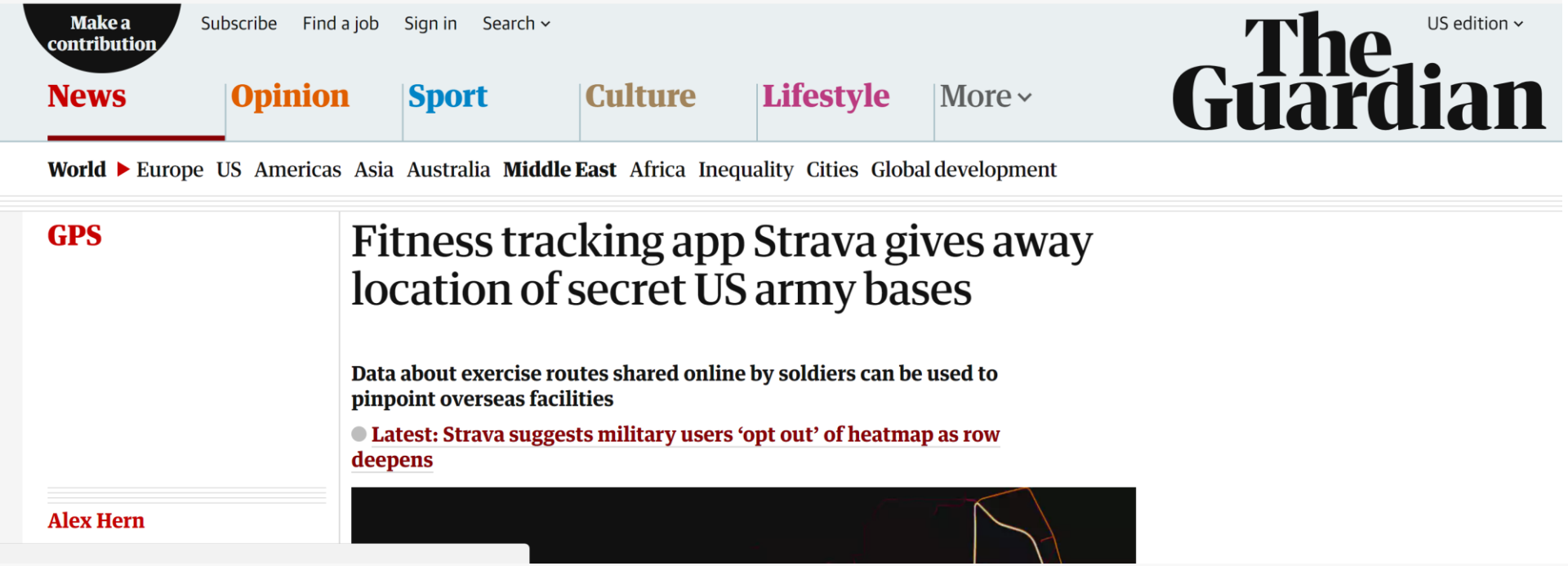
Challenges: Privacy vs. Utility

- Check-in risks: Stalking, GeoSlavery, Others know that you are not home, etc
- Ex: Girls Around me App (3/2012)

The Girls of Girls Around Me. It's doubtful any of these girls even know they are being tracked. Their names and locations have been obscured for privacy reasons. (Source: Cult of Mac, March 30, 2012)

Challenges: Security vs. Utility

Challenges: Security vs. Utility



Make a contribution Subscribe Find a job Sign in Search US edition

The Guardian

News Opinion Sport Culture Lifestyle More

World Europe US Americas Asia Australia Middle East Africa Inequality Cities Global development

GPS

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users 'opt out' of heatmap as row deepens

Alex Hern

Location-based threats: How cybercriminals target you based on where you live

Corporate · Network · Security Tips · SophosLabs · Cryptowall · Geomalware · Locky · Phishing · Ransomware · Sophos Home · Spam · TorrentLocker

Challenges: Security vs. Utility

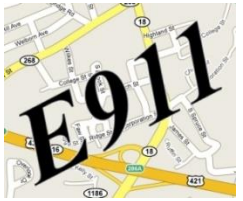
- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?

Challenges: Security vs. Utility

- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?
- Involved groups:
  - Civil Society
  - Economic Entities
  - Public Safety
  - Policy Makers

Challenges: Security vs. Utility

- Important questions:
  - Who gets my data?
  - Who do they give it to?
  - What promises do I get?
- Involved groups:
  - Civil Society
  - Economic Entities
  - Public Safety
  - Policy Makers
- Agreements and disagreements
  - Agreements: E911, emergency alerts
  - Controversial: traffic monitoring



Spatial beyond GeoSpatial

- Examples:
  - Human bodies
  - VLSI
  - Universe

Spatial beyond GeoSpatial

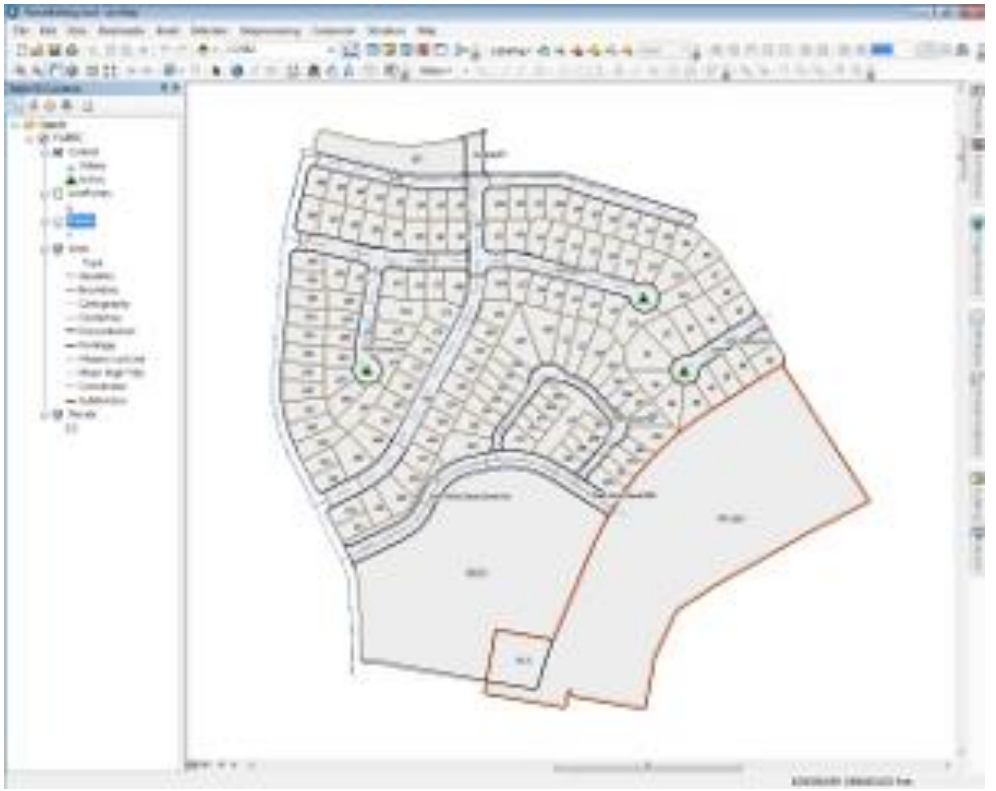
- Examples:
  - Human bodies
  - VLSI chips and boards
  - Universe
  - Indoor and virtual spaces
- Challenges:
  - What are the reference system?
    - On Mars? Outside Milkyway galaxy? In augmented reality spaces?
    - Is it one for all humans? Or personalized?
  - Accuracy
  - 3D+ scalability

Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

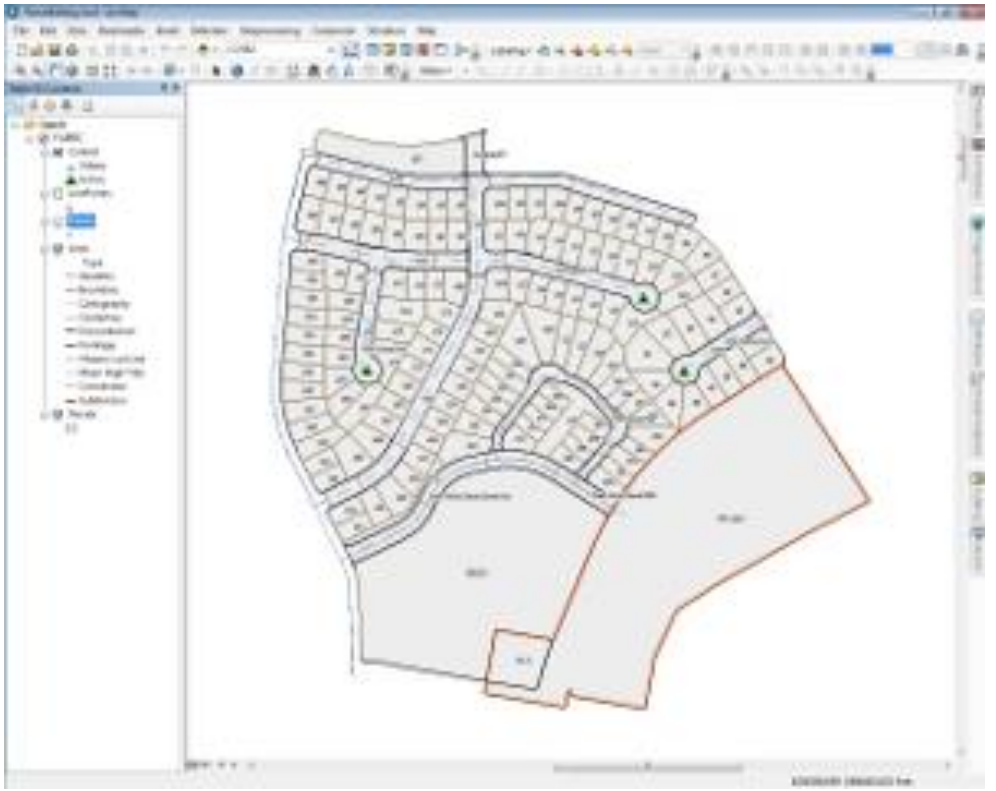
Geographic Information Systems (GIS)

- Software packages for working with maps and geographic information.
  - Creating and using maps
  - Compiling geographic data
  - Analyzing mapped info
  - Sharing and discovering geographic information



Geographic Information Systems (GIS)

- Software packages for working with maps and geographic information.
  - Creating and using maps
  - Compiling geographic data
  - Analyzing mapped info
  - Sharing and discovering geographic information



How different GIS from SDBMS?

How different GIS from SDBMS?

- GIS uses SDBMS to store, search, and query spatial data

How different GIS from SDBMS?

- GIS uses SDBMS to store, search, and query spatial data
- GIS is a software application, SDBMS is a data management system

How different GIS from SDBMS?

- GIS uses SDBMS to store, search, and query spatial data
- GIS is a software application, SDBMS is a data management system
- GIS used to visualize and analyze spatial data
  - Rich high-level analysis
- SDBMS used to store, index, and query spatial data efficiently
  - Efficient and scalable fundamental querying and data management operations

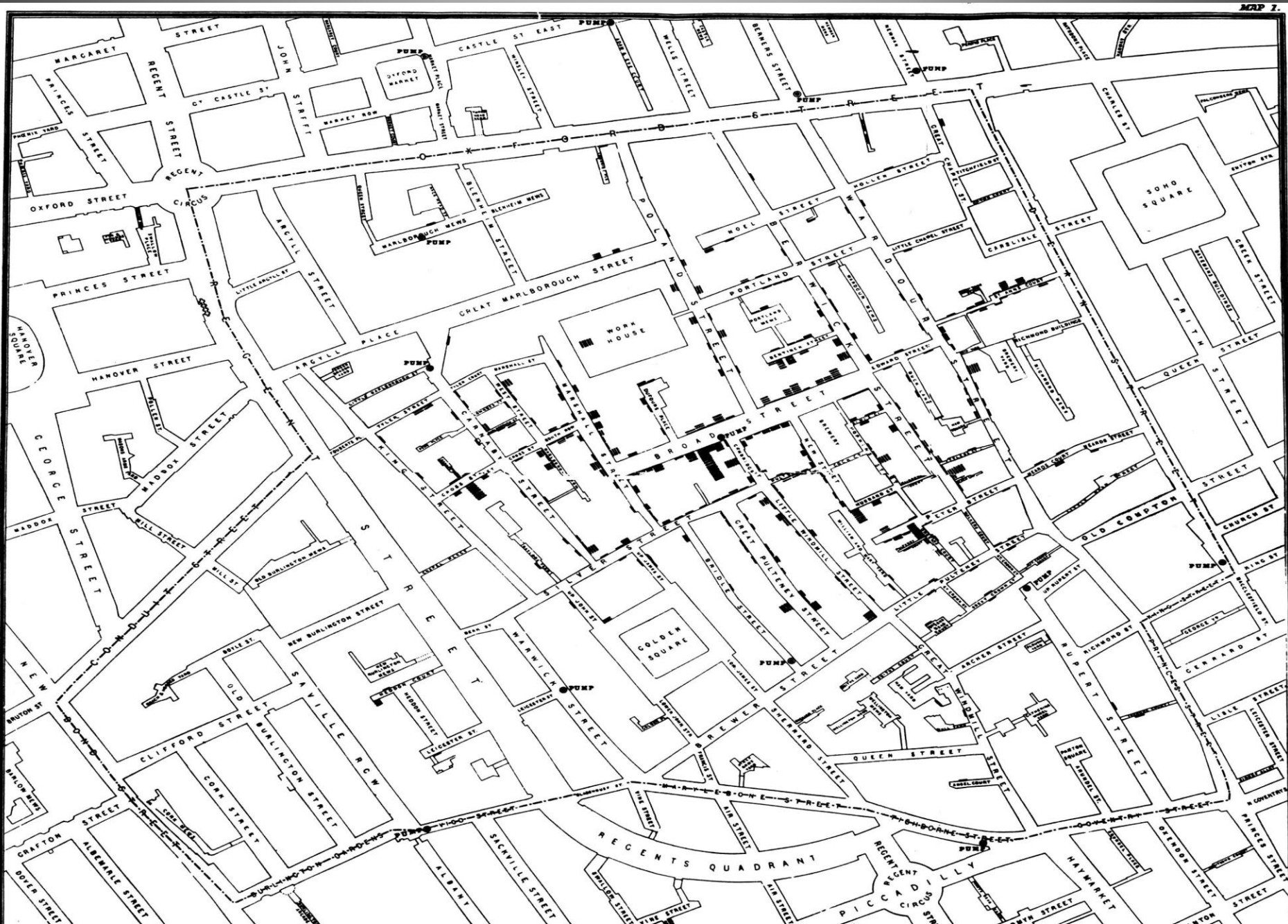
How different GIS from SDBMS?

- GIS uses SDBMS to store, search, and query spatial data
- GIS is a software application, SDBMS is a data management system
- GIS used to visualize and analyze spatial data
  - Rich high-level analysis
- SDBMS used to store, index, and query spatial data efficiently
  - Efficient and scalable fundamental querying and data management operations
- SDBMS can be used by applications other than GIS
  - Astronomy, location-based services, brain informatics, etc

Major technologies and areas (past, present, & future)

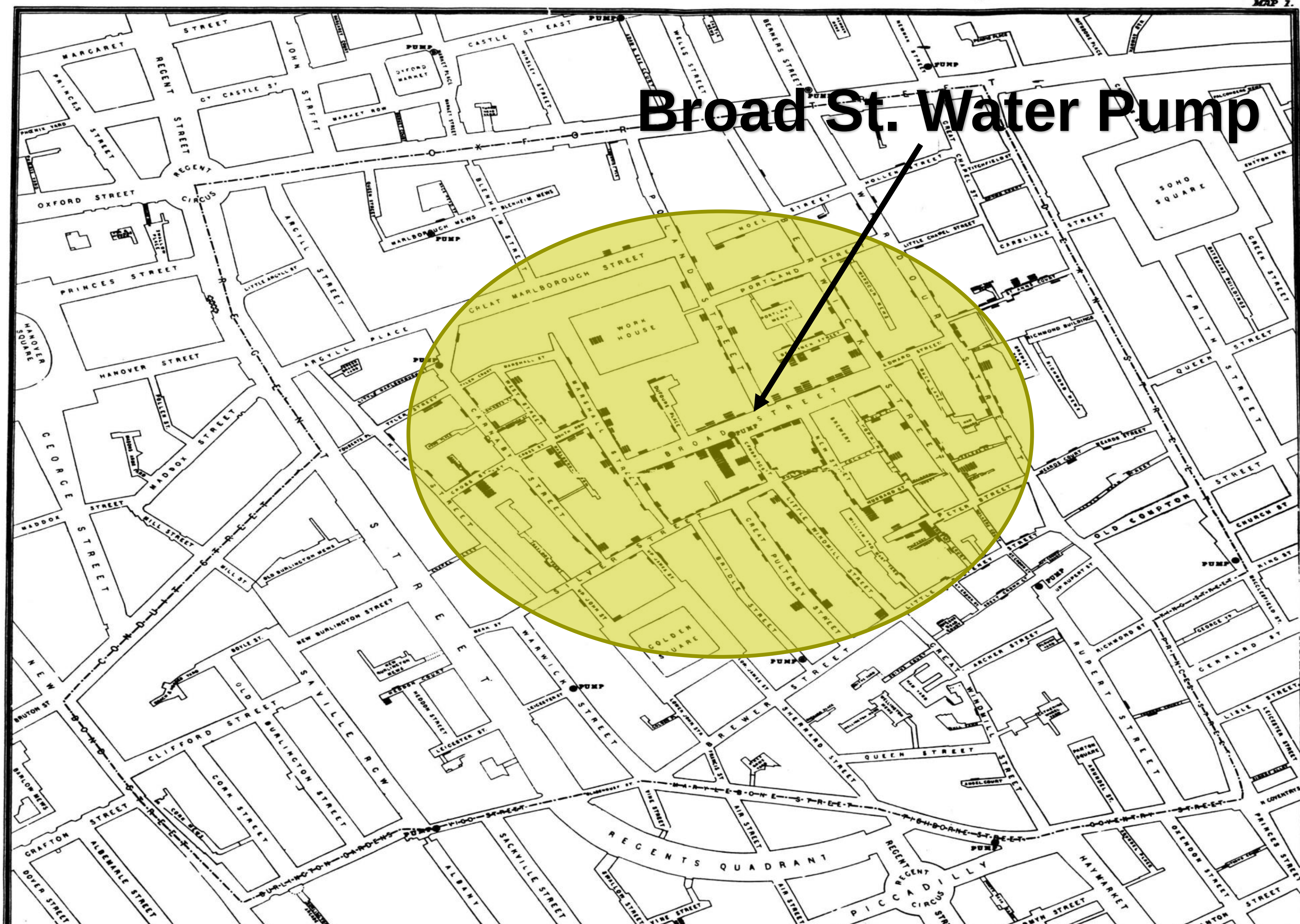
- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

Cholera cases in the London epidemic of 1854



Cholera cases in the London epidemic of 1854

Cholera cases in the London epidemic of 1854



34 COMMENTS

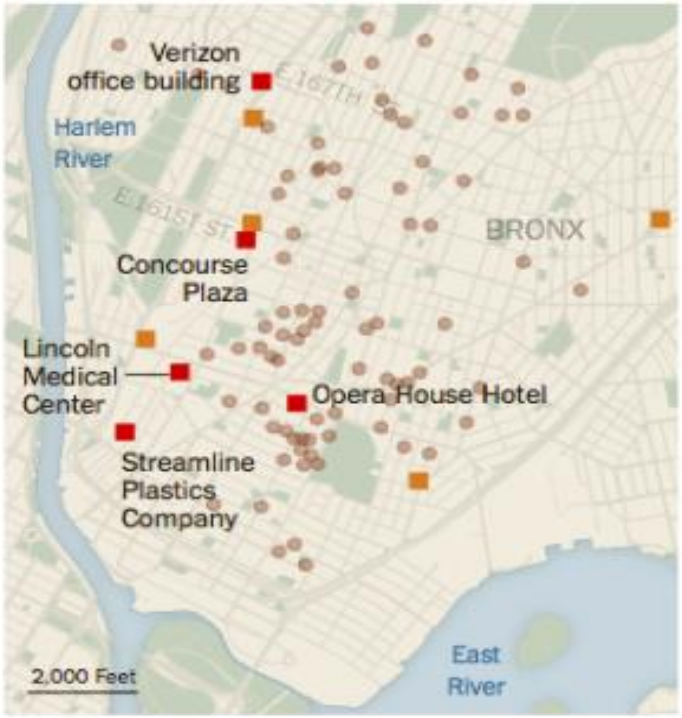
Hotel That Enlivened the Bronx Is Now a 'Hot Spot' for Legionnaires'

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

34 COMMENTS

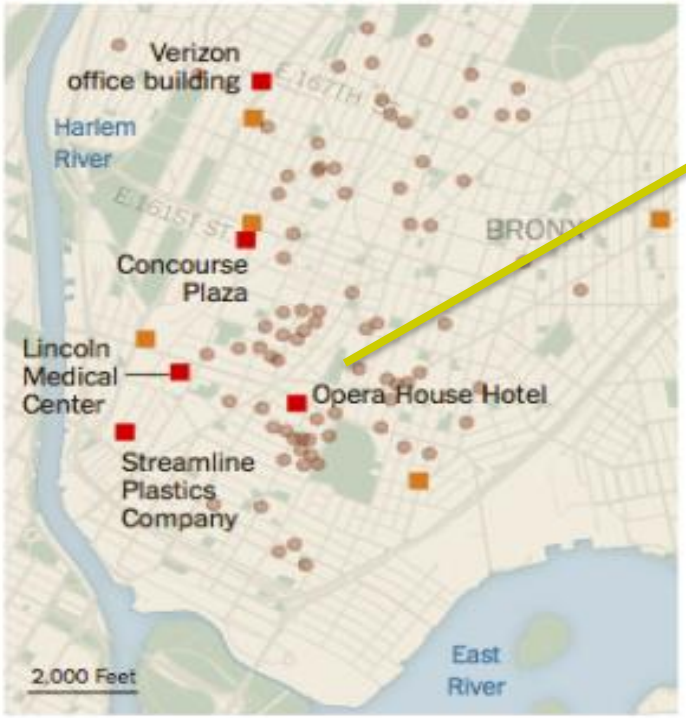
Hotel That Enlivened the Bronx Is Now a 'Hot Spot' for Legionnaires'

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'



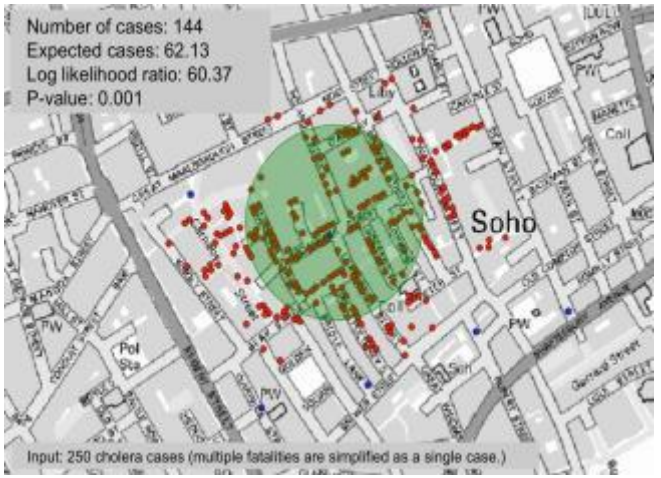
Source: New York Mayor's Office

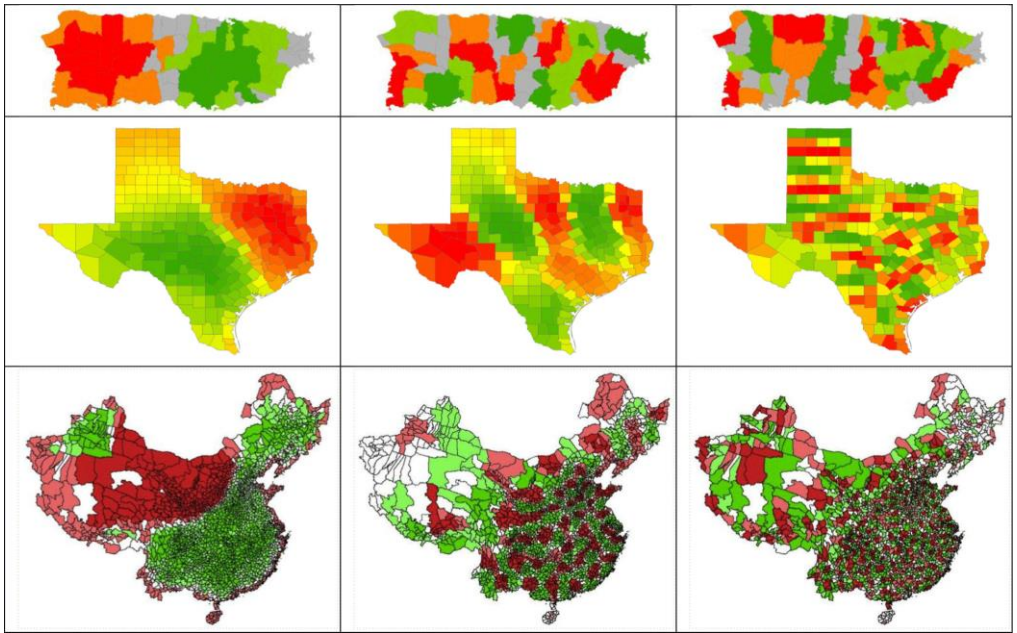
By The New York Times

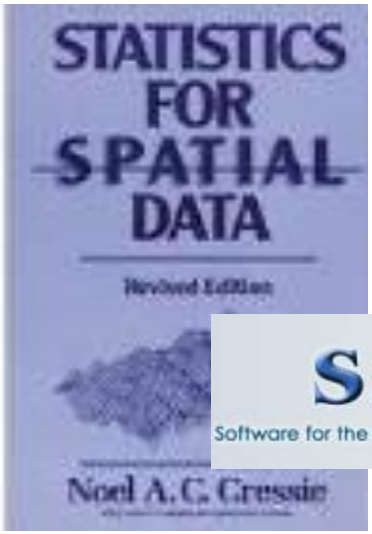
The Opera House Hotel is at the center of the outbreak. Edwin J. Torres for The New York Times

Spatial Statistics

- In the spatial space, statistical independence assumptions do not always hold
- Spatial Statistics
  - Hot spot detection
  - Spatial auto-correlation
  - Spatial-constrained clusters
  - Spatial uncertainty, confidence, etc



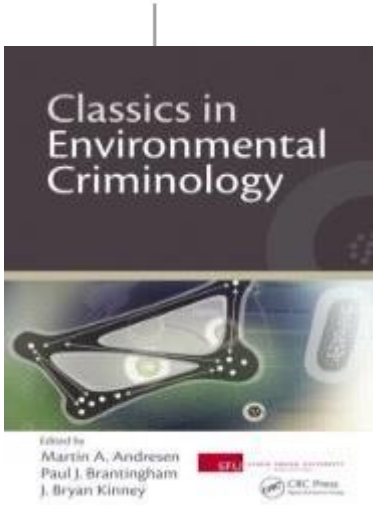




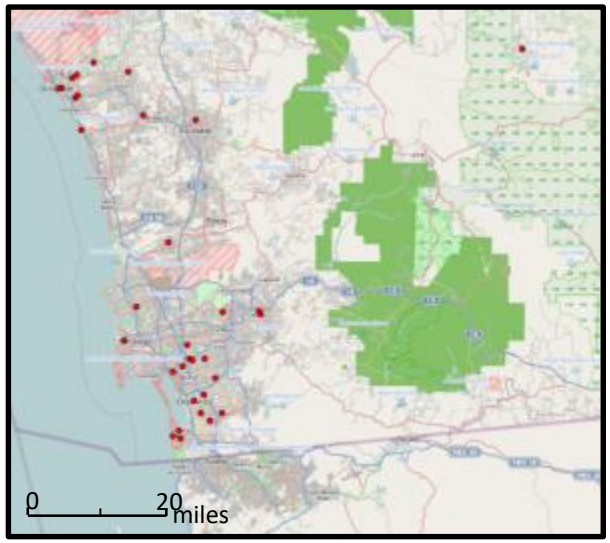
SaTScan™
Software for the spatial, temporal, and space-time scan statistics

Detecting Spatial Patterns

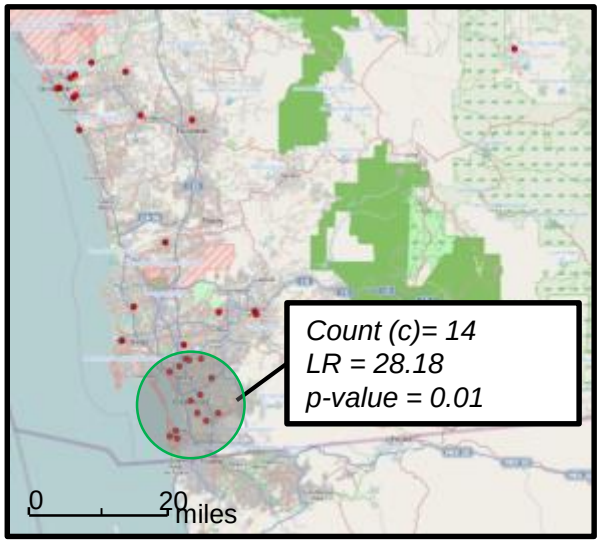
- Arson crimes in San Diego in 2013
  - Total 33 cases (red dots on the map)
  - Activity Area is appr. 3000 sq. miles.
- Arsonist caught in top green ring[2]



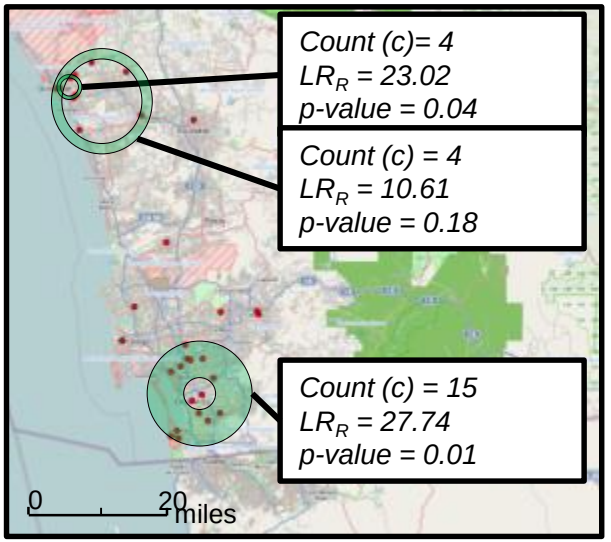
Input



SaTScan output



Significant Ring Detection

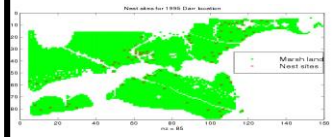


Green: Rings with LR >10 & p-value < 0.20

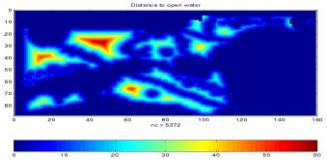
(1) http://www.sandiego.gov/police/services/statistics/index.shtml
(2) http://www.nbcsandiego.com/news/local/Suspected-Arson-Grass-Fires-Oceanside-Mesa-Drive-Foussat-Road-218226321.html
(3) Ring-Shaped Hot-Spot Detection: A Summary of Results, IEEE Intl. Conf. on Data Mining, 2014.

Location Prediction: nesting sites

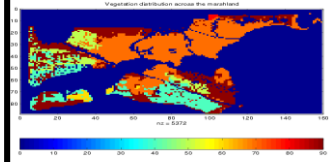
Nest locations



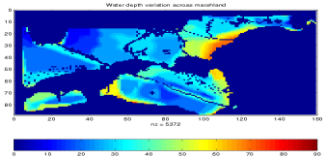
Distance to open water



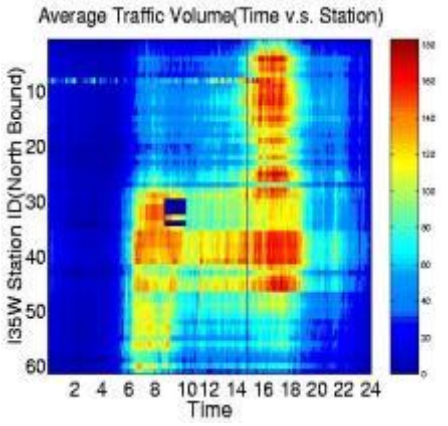
Vegetation durability



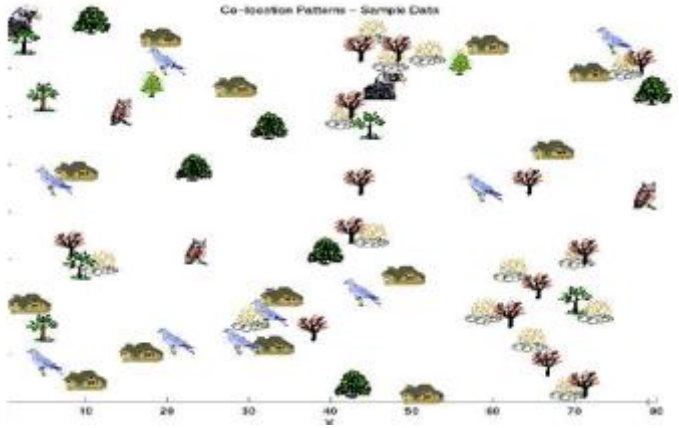
Water depth



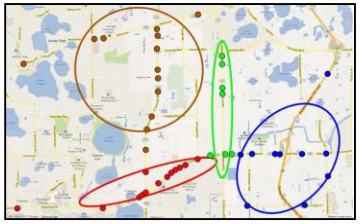
Spatial outliers: sensor (#9) on I-35

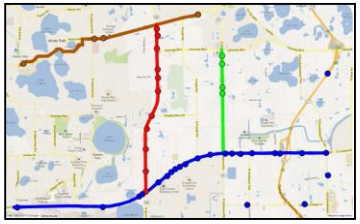


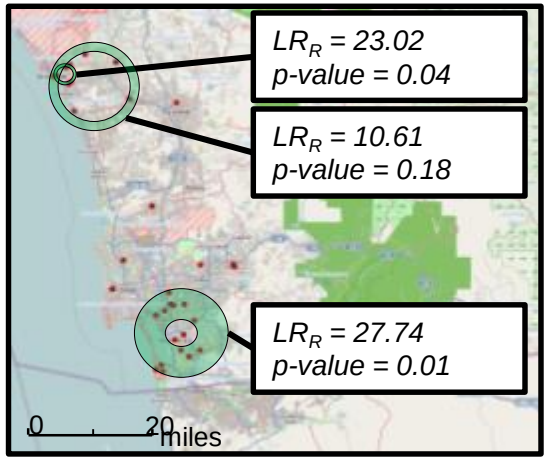
Co-location Patterns



Spatial Concept Aware Summarization





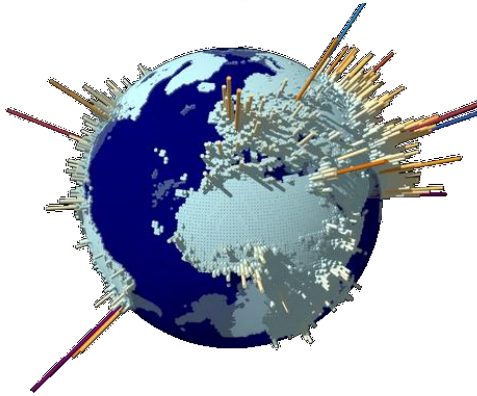


Major technologies and areas (past, present, & future)

- GPS
- Location Based Services
- Spatial Data Management Systems
- Geographic Information Systems
- Spatial Predictive Analysis (Spatial Statistics, or Spatial Data Mining)
- Virtual Globes and VGI (or CGI)

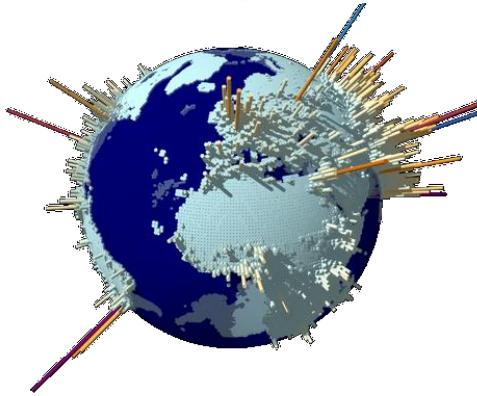
Virtual Globes and VGI (or CGI)

- LBS accessibility
- Visualization
- Volunteering (or Crowdsourcing) geo information
- Education



Virtual Globes and VGI (or CGI)

- LBS accessibility
- Visualization
- Volunteering (or Crowdsourcing) geo information
- Education



Virtual Globes in GIS Education

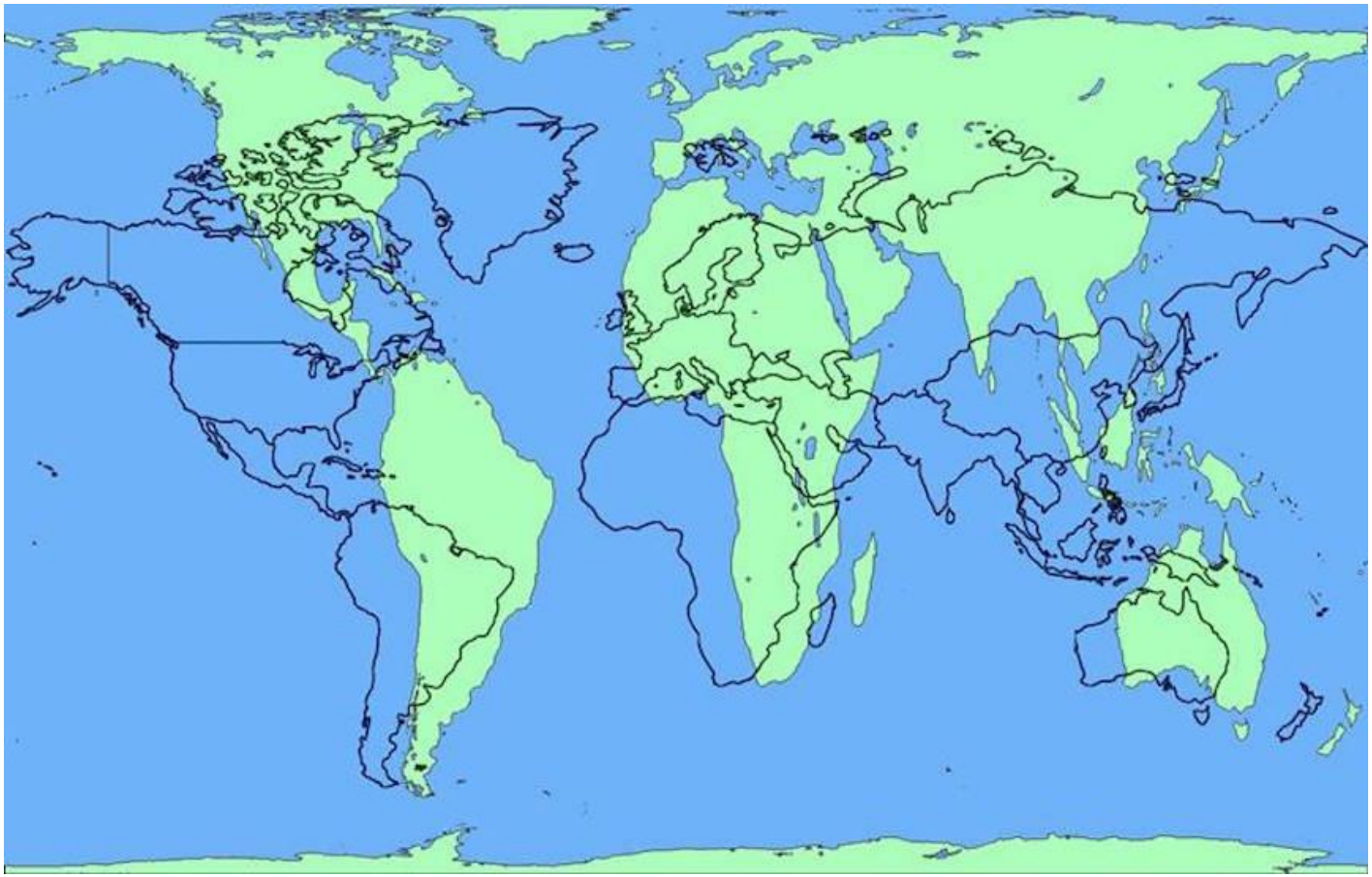
- Coursera MOOC: From GPS and Google Earth to Spatial Computing
  - 21,844 students from 182 countries (Fall 2014)
  - 8 modules, 60 short videos, in-video quizzes, interactive examinations, …
  - 3 Tracks: curious, concepts, technical

Map Orientation and Projections

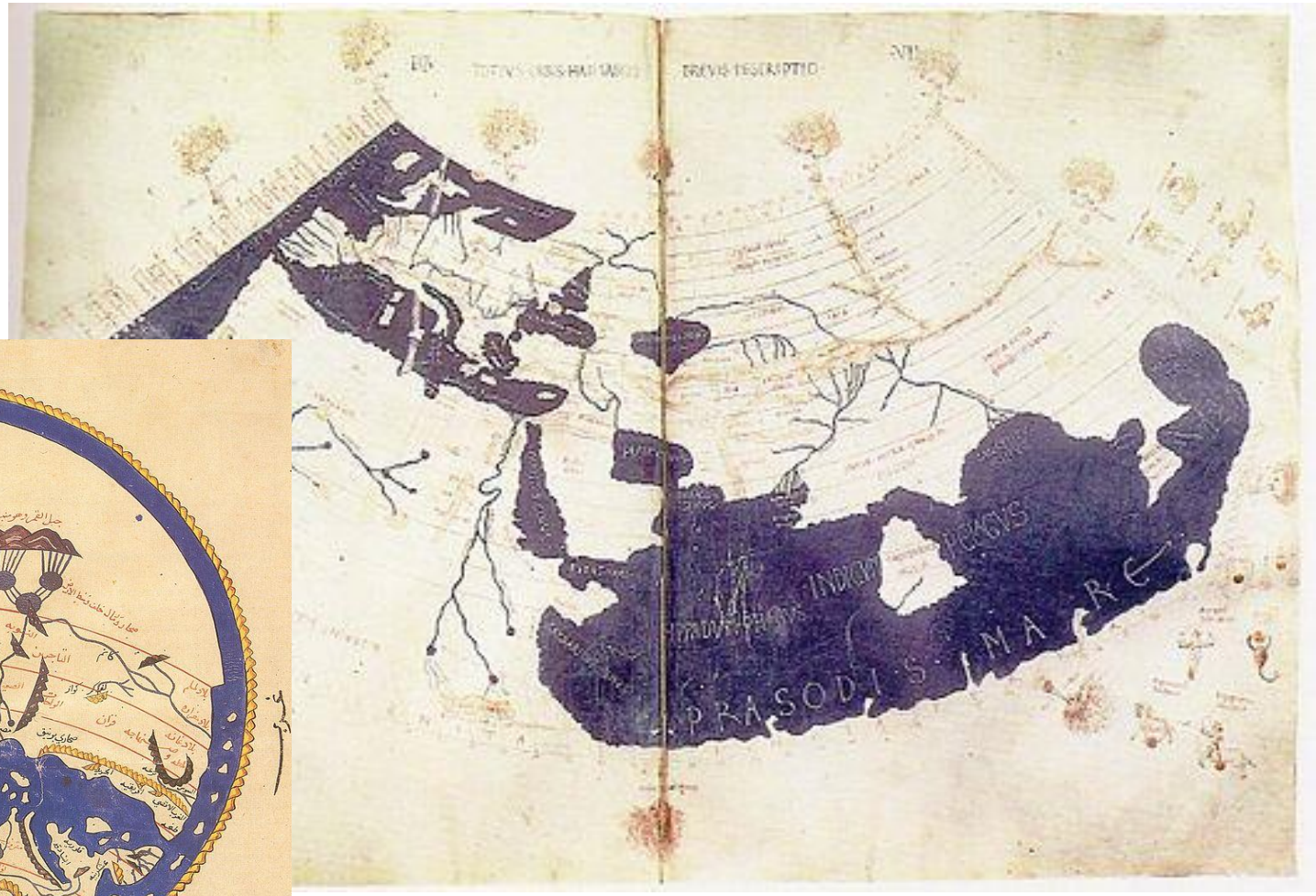
- Mapping a 3D globe on a flat 2D plane
  - https://www.youtube.com/watch?v=kIID5FDi2JQ

Map Orientation and Projections

- Mapping a 3D globe on a flat 2D plane
  - https://www.youtube.com/watch?v=kIID5FDi2JQ



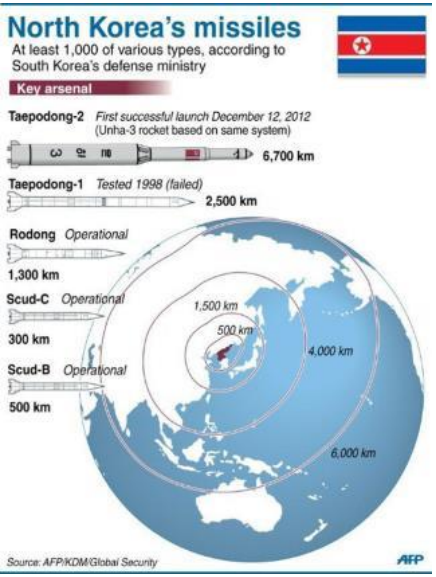
Map Orientation and Projections



Map Orientation and Projections

Map Orientation and Projections

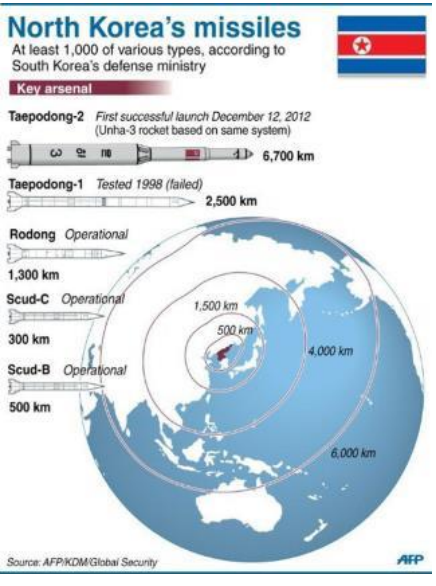
The Economist



Map Orientation and Projections

The Economist

Original



Correction