

The logo for the University of California, Riverside (UCR), featuring the letters 'UCR' in a bold, sans-serif font. The 'UC' is orange and the 'R' is blue with a white sunburst symbol on its right side.

UCR

A smaller version of the UCR logo, consisting of the letters 'UCR' in orange and blue with a white sunburst symbol on the 'R', positioned in the top right corner of the slide.

UCR

The main title of the slide, 'Spatial Computing', is written in a large, bold, sans-serif font. 'Spatial' is in black and 'Computing' is in white. The text is overlaid on a background featuring a stylized globe with a grid of latitude and longitude lines, and a pattern of blue and green dots on the left side.

**Spatial  
Computing**

# Geovisualization

Amr Magdy

Computer Science and Engineering Department

Center for Geospatial Sciences

University of California, Riverside

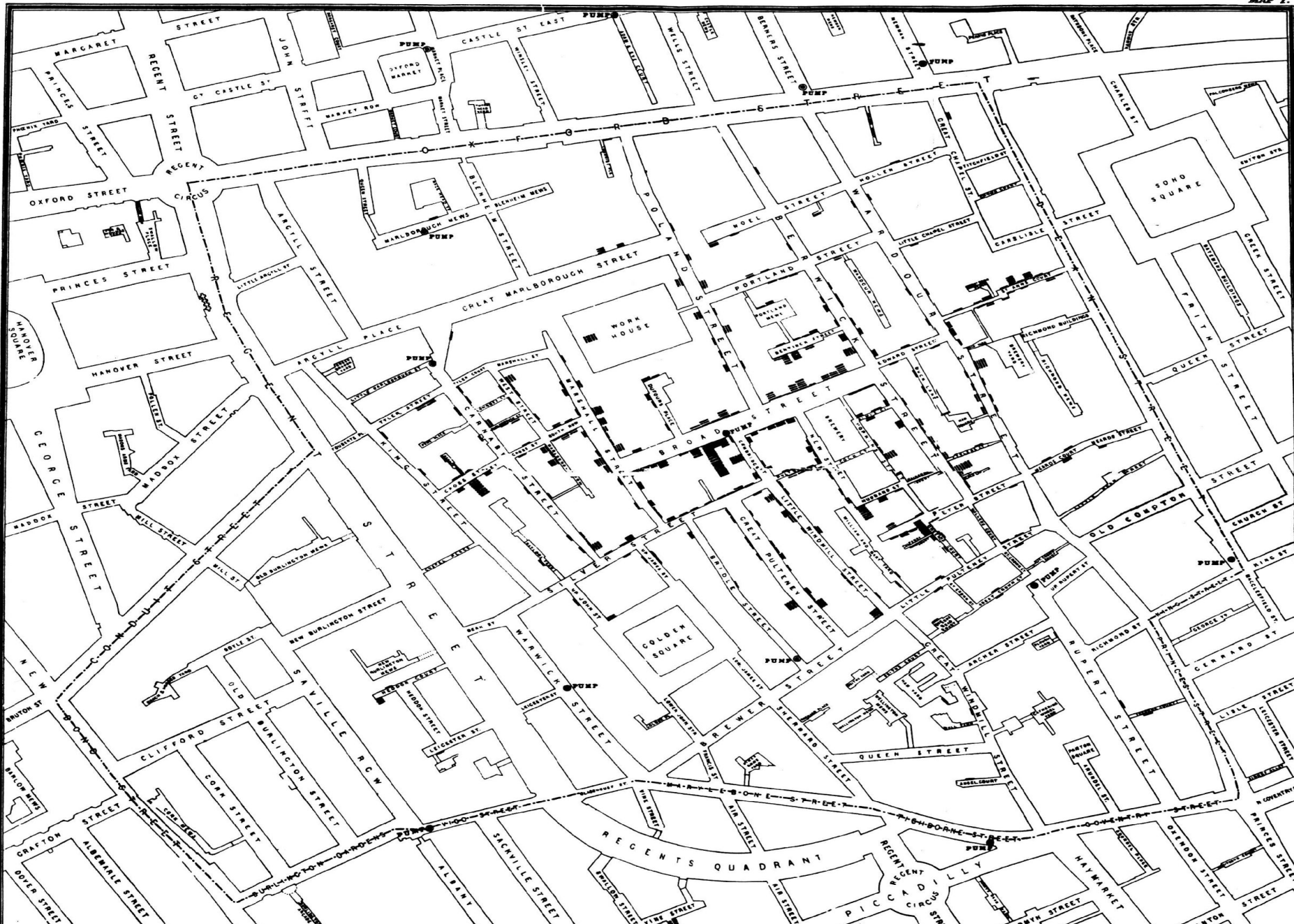
[www.cs.ucr.edu/~amr/](http://www.cs.ucr.edu/~amr/)

# Visual Perception

- Learning Styles & Personality Types: Visual, Auditory, Kinesthetic



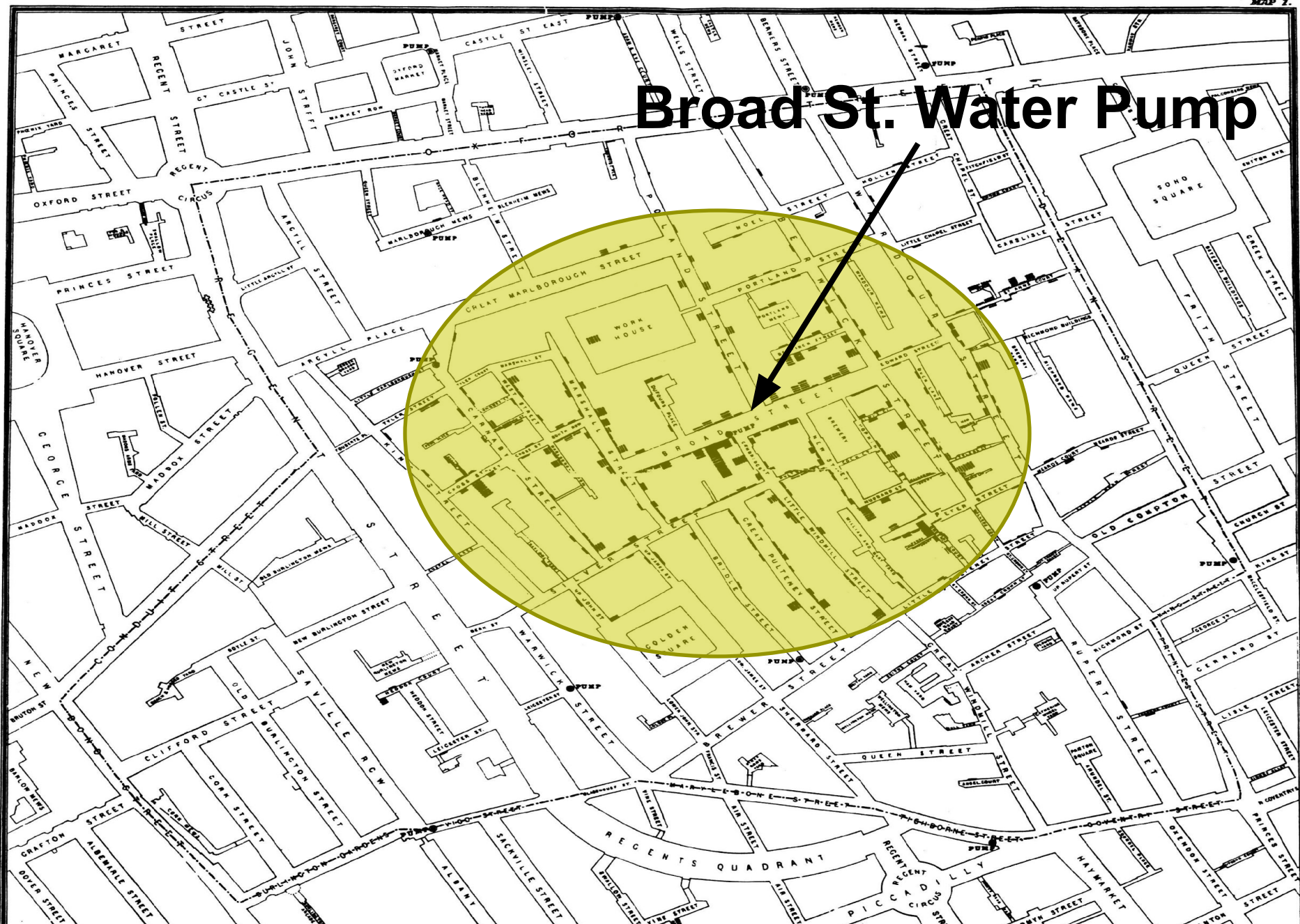
# Cholera cases in the East London epidemic of 1854



# Cholera cases in the East London epidemic of 1854



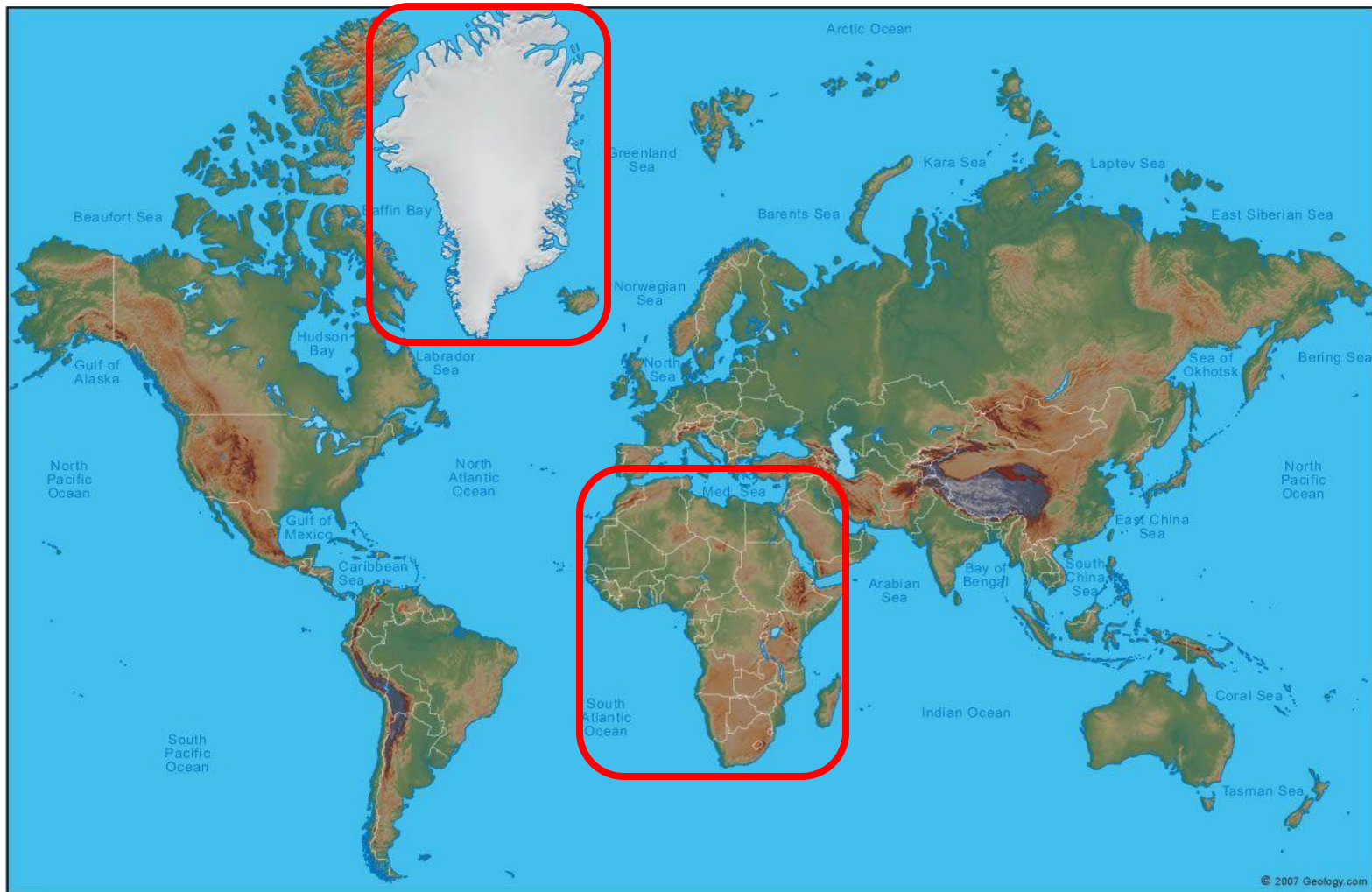
# Cholera cases in the London epidemic of 1854



**Broad St. Water Pump**

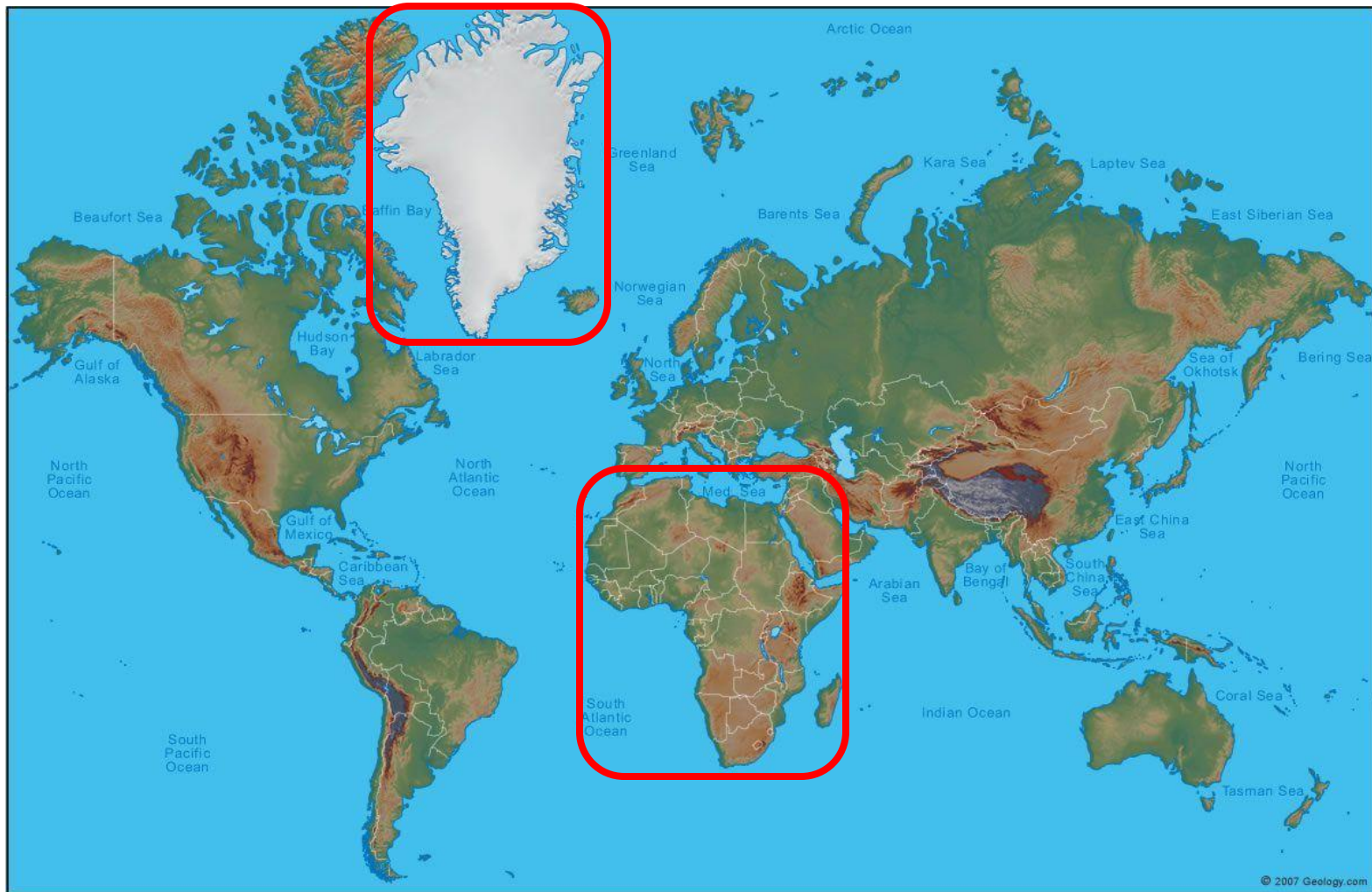
# Geo-Visualization

- What is the ratio between areas of Africa and Greenland?



# Geo-Visualization

- What is the ratio between areas of Africa and Greenland? 14:1





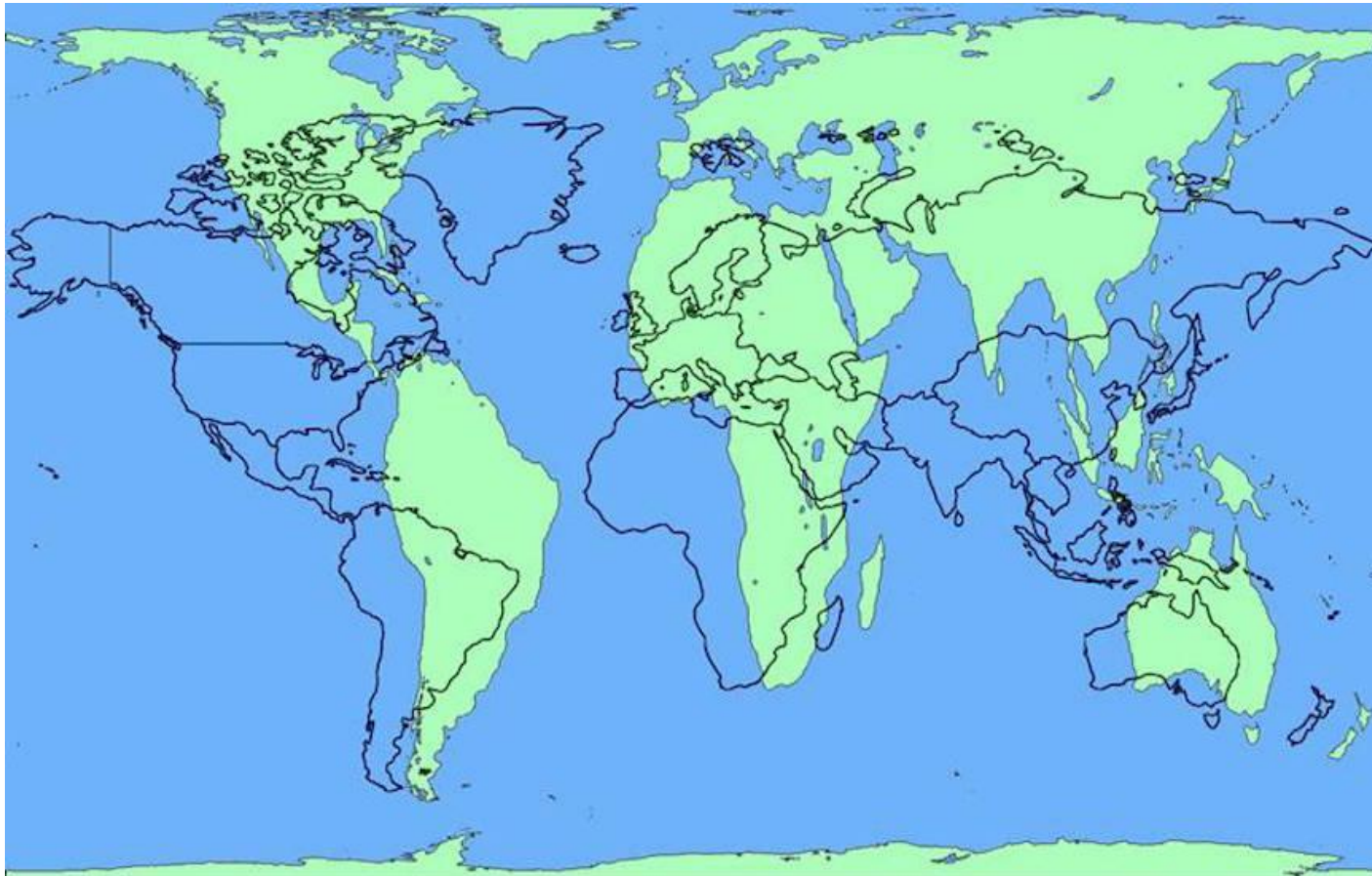
# Map Orientation and Projections

- Mapping a 3D globe on a flat 2D plane
  - Why all world maps are wrong?
    - <https://www.youtube.com/watch?v=kIID5FDi2JQ>

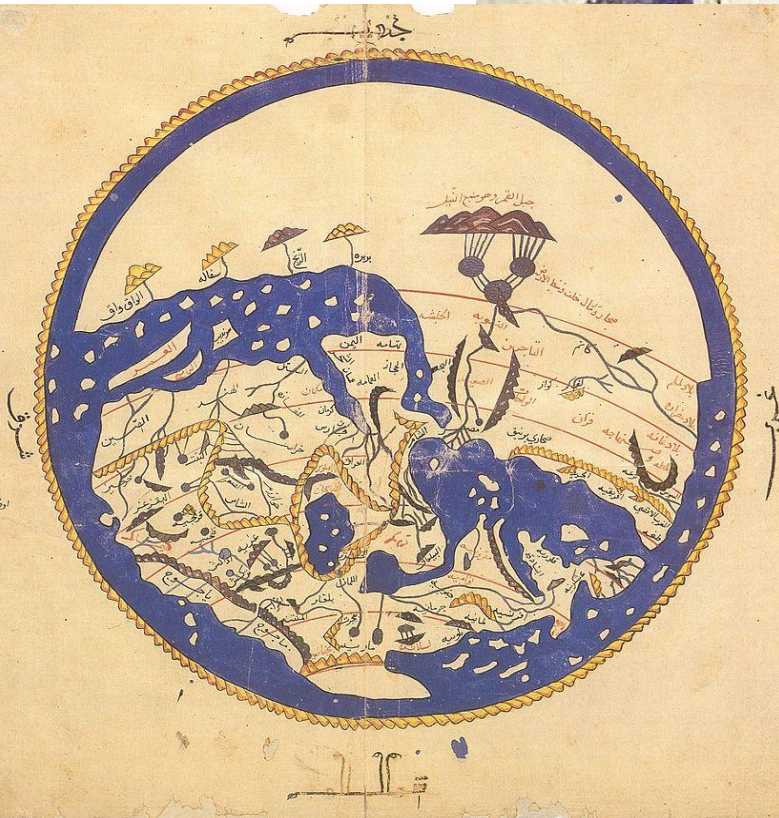
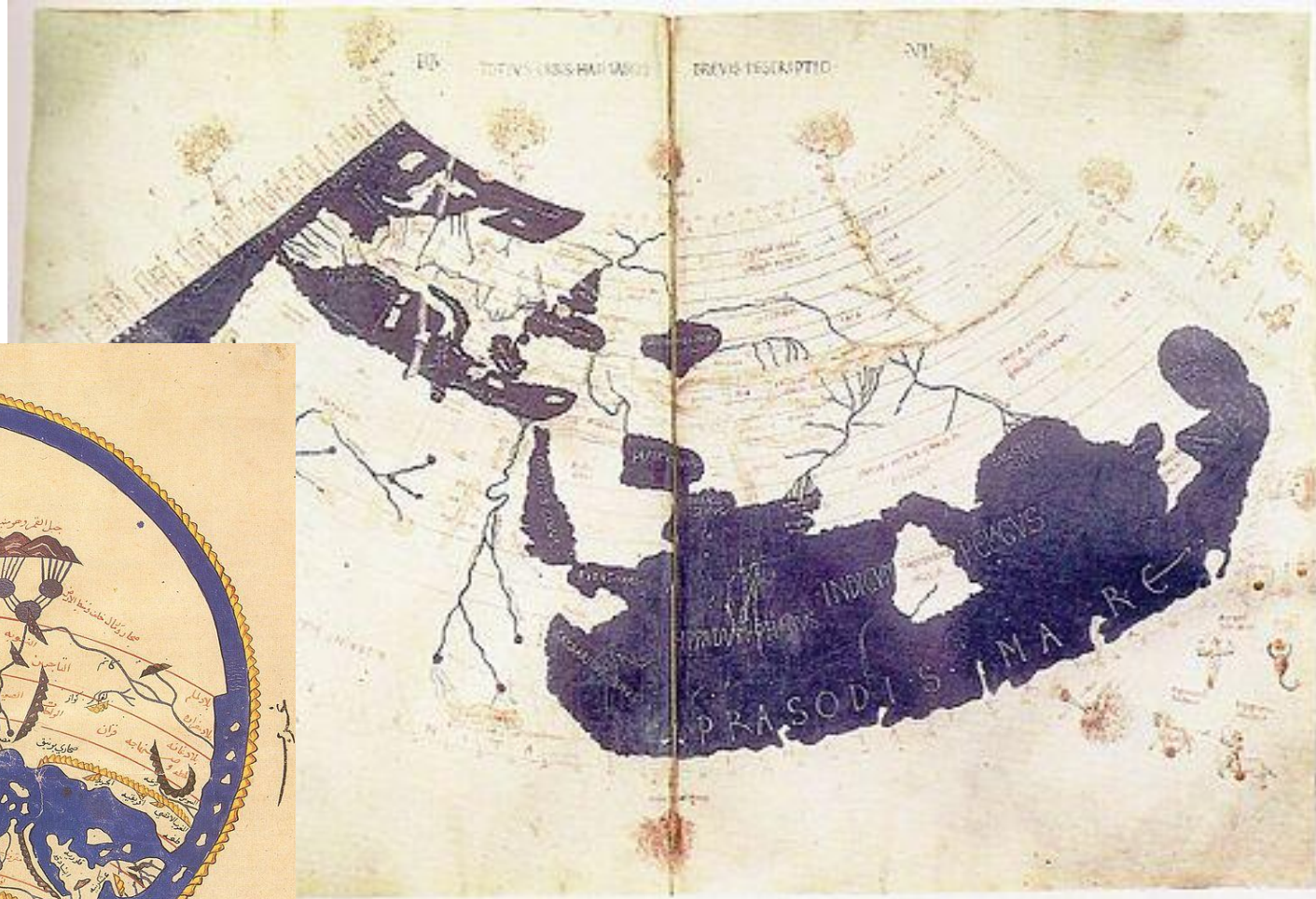


# Map Orientation and Projections

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



# Map Orientation and Projections



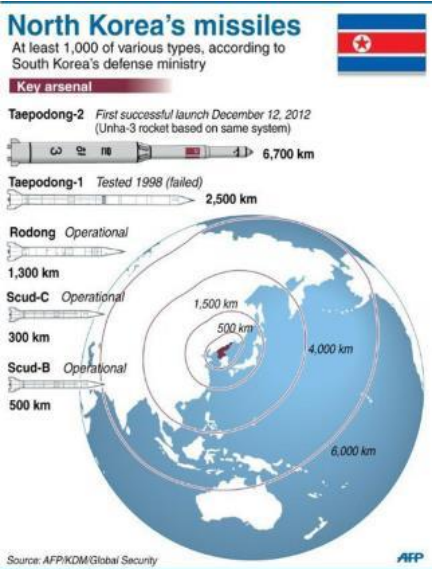
# Map Orientation and Projections



# Map Orientation and Projections



The Economist



# Map Orientation and Projections



The Economist

Correction

Original



## North Korea's missiles

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



### Key arsenal

**Taepodong-2** First successful launch December 12, 2012 (Unha-3 rocket based on same system)  
6,700 km



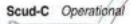
**Taepodong-1** Tested 1998 (failed)  
2,500 km



**Rodong** Operational  
1,300 km



**Scud-C** Operational  
300 km



**Scud-B** Operational  
500 km



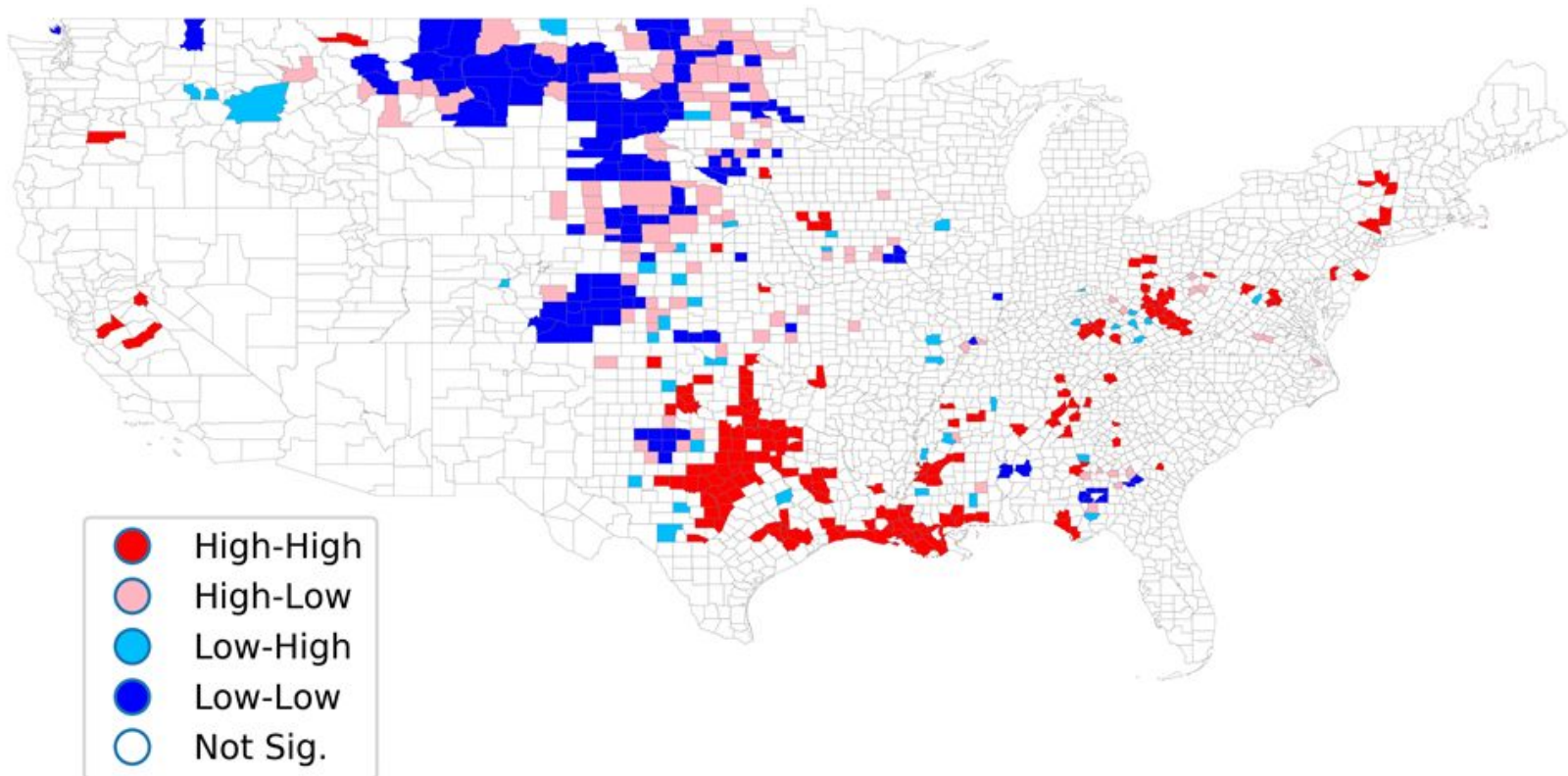
# Why?



- Why visualization?
  - Get insights
  - Come up with hypotheses
  - Detect the expected, and discover the unexpected ®

# Why?

- Why visualization?
  - Get insights
  - Come up with hypotheses
  - Detect the expected, and discover the unexpected ®





# Applications

- Mapping
  - With all map applications throughout history
- Decision making
  - E.g., disease outbreaks, crimes, etc
- Real-time monitoring
  - E.g., traffic, security, etc
- Scientific analysis
  - E.g., climate change, vegetation analysis, etc
- ...

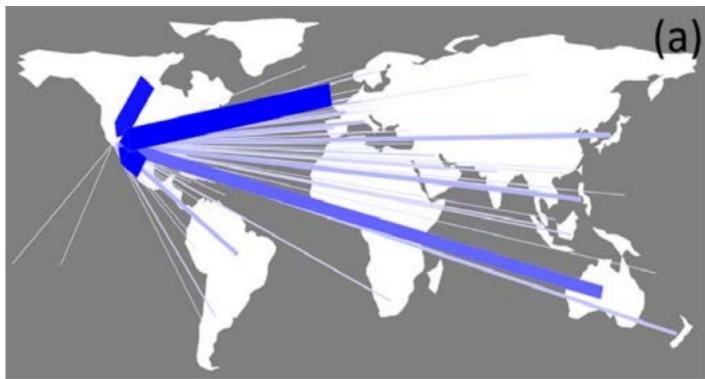
# Geo-visualization Element



- Three elements
  - Data: what to visualize?
  - Location: where to put data?
  - Visualization scheme: how to visualize?

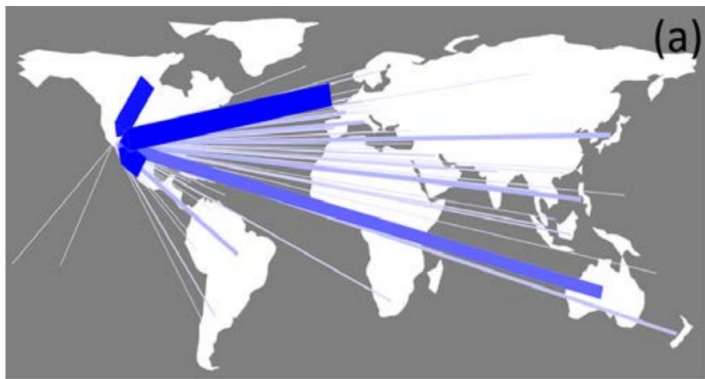
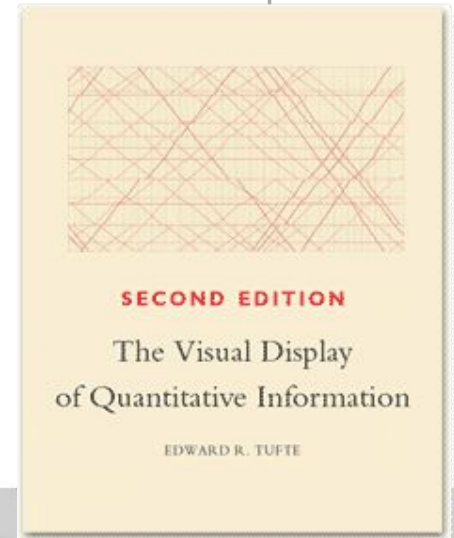
# Geo-visualization Element

- Three elements
  - Data: what to visualize?
  - Location: where to put data?
  - Visualization scheme: how to visualize?



# Geo-visualization Element

- Three elements
  - Data: what to visualize?
  - Location: where to put data?
  - Visualization scheme: how to visualize?

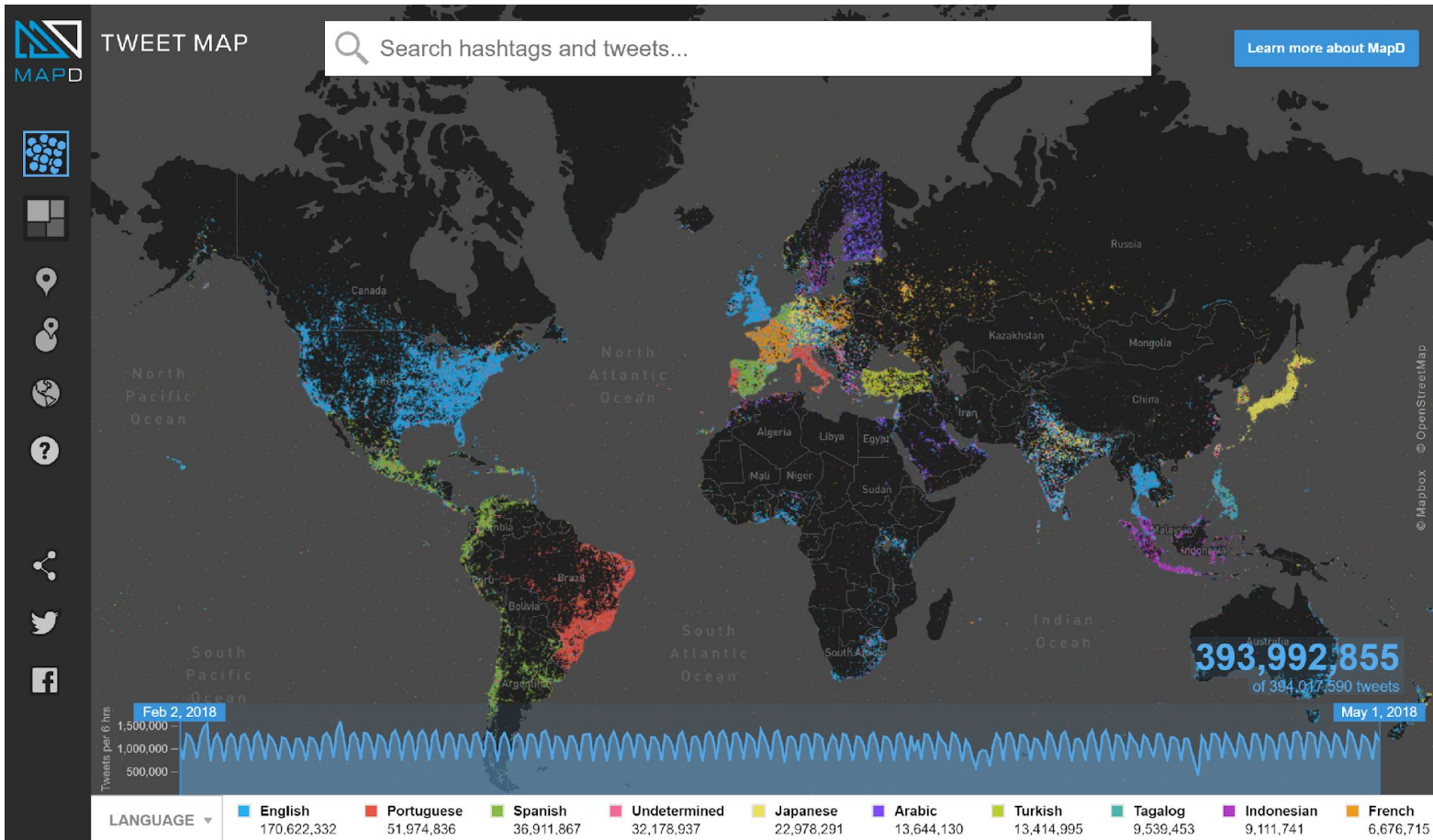


# Interactive Maps



- MapD interactive demos

- Tweet map: <https://www.mapd.com/demos/tweetmap/>

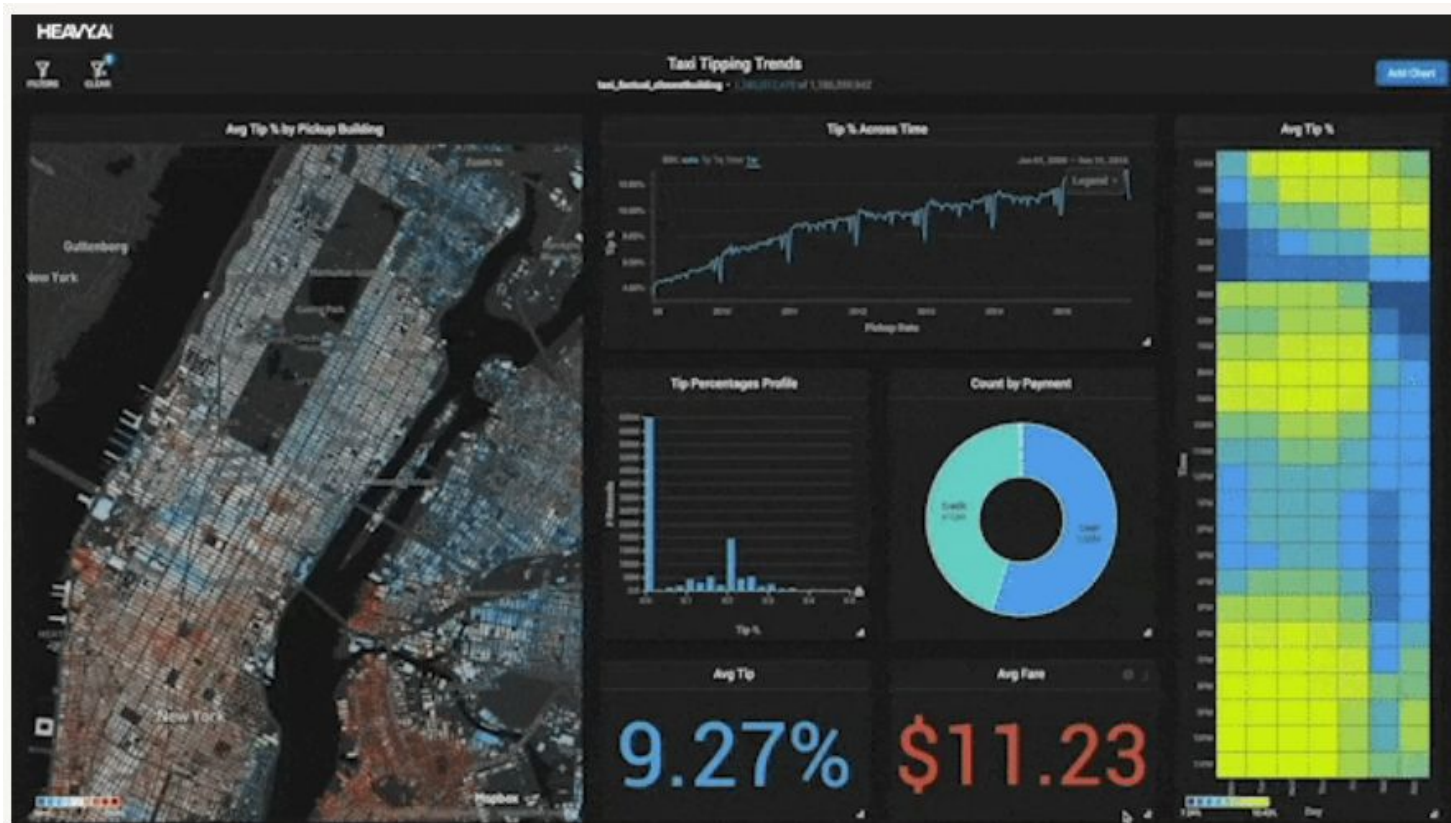


## TOP HASHTAGS TWEETS

#heartwards	595,879
#bestfanarmy	462,436
#premiosmtvmiaw	373,845
#trndnl	315,228
#bbb18	303,674
#btsarmy	276,614
#nowplaying	228,440
#exol	196,345
#repost	178,637
#photo	176,390
#love	

# Interactive Maps

- Heavy Ai interactive demos
  - NYC Taxi:  
<https://demo-taxis.heavy.ai/>



# Interactive Maps

- Pan and Zoom (in interactive views)
  - Pan: change your data focus on same spatial view level
  - Zoom: change your spatial view level

# Interactive Maps

- Pan and Zoom (in interactive views)
  - Pan: change your data focus on same spatial view level
  - Zoom: change your spatial view level
- Linking and Brushing (in multiple views)
  - Linking: highlight certain part of data in all views
  - Brushing: dynamic linking (linking + panning)
  - This happens when you have multiple distinct views, e.g., a map, a table, and a graph, or a set of temporally partitioned views



# Interactive Maps

- Pan and Zoom (in interactive views)
  - Pan: change your data focus on same spatial view level
  - Zoom: change your spatial view level
- Linking and Brushing (in multiple views)
  - Linking: highlight certain part of data in all views
  - Brushing: dynamic linking (linking + panning)
  - This happens when you have multiple distinct views, e.g., a map, a table, and a graph, or a set of temporally partitioned views
- Specification of interactive visualization
  - 200 ms response time (controversial)

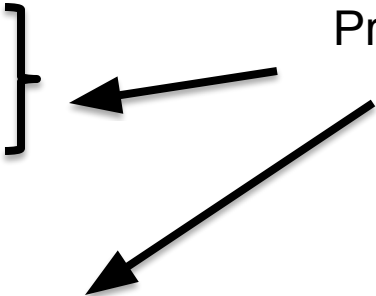


# Big Spatial Data Visualization



- New challenges come with big volume data
  - How to put data on the map?
  - How to aggregate large data?
  - How to process large data?

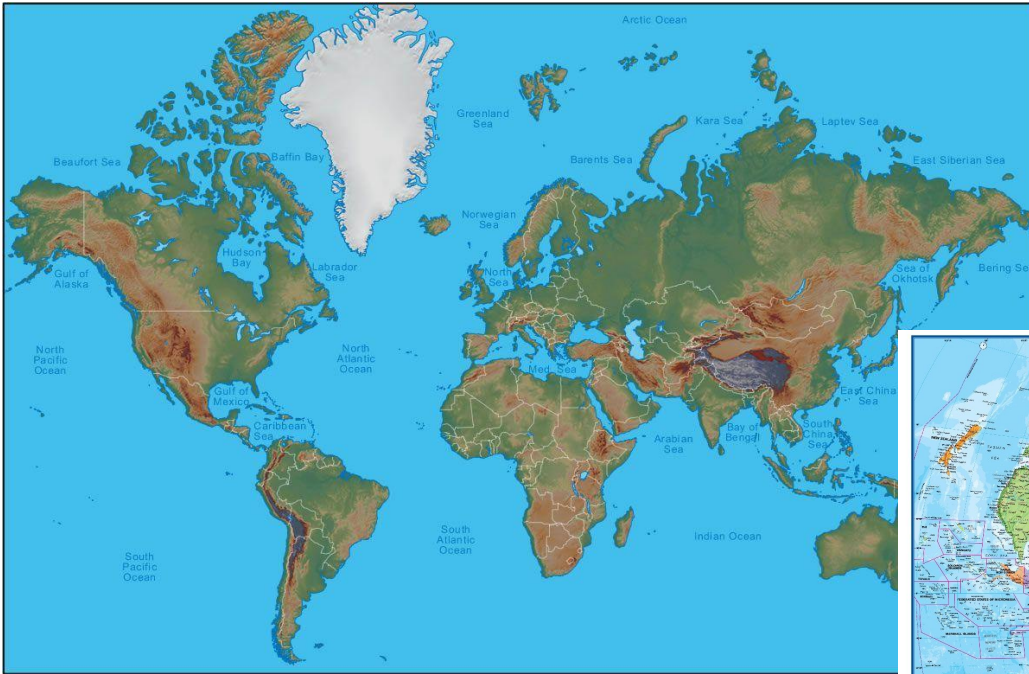
# Big Spatial Data Visualization

- New challenges come with big volume data
    - How to put data on the map?
    - How to aggregate large data?
    - How to process large data?
- 

Pre-processing (preparing)  
data for visualization
- High velocity
    - High velocity data visualization exploits pre-materialization
    - Still active research is on-going

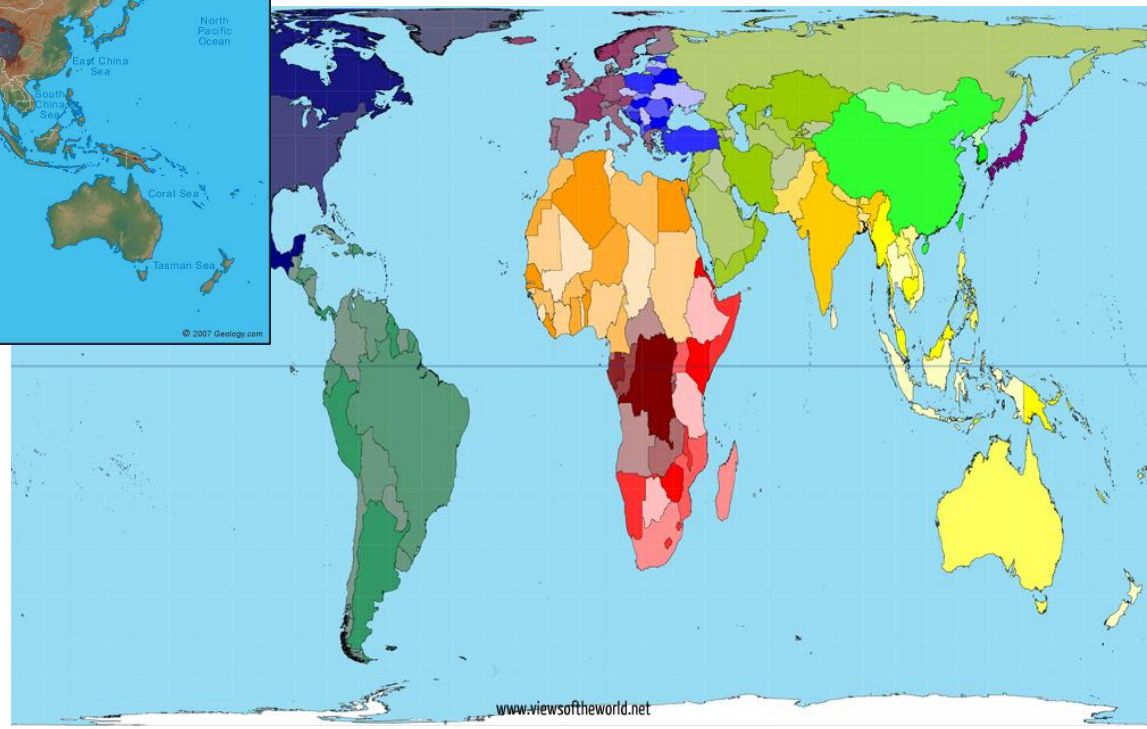
# Designing an Effective Visualization

- Need to take human perception into account (orientation)



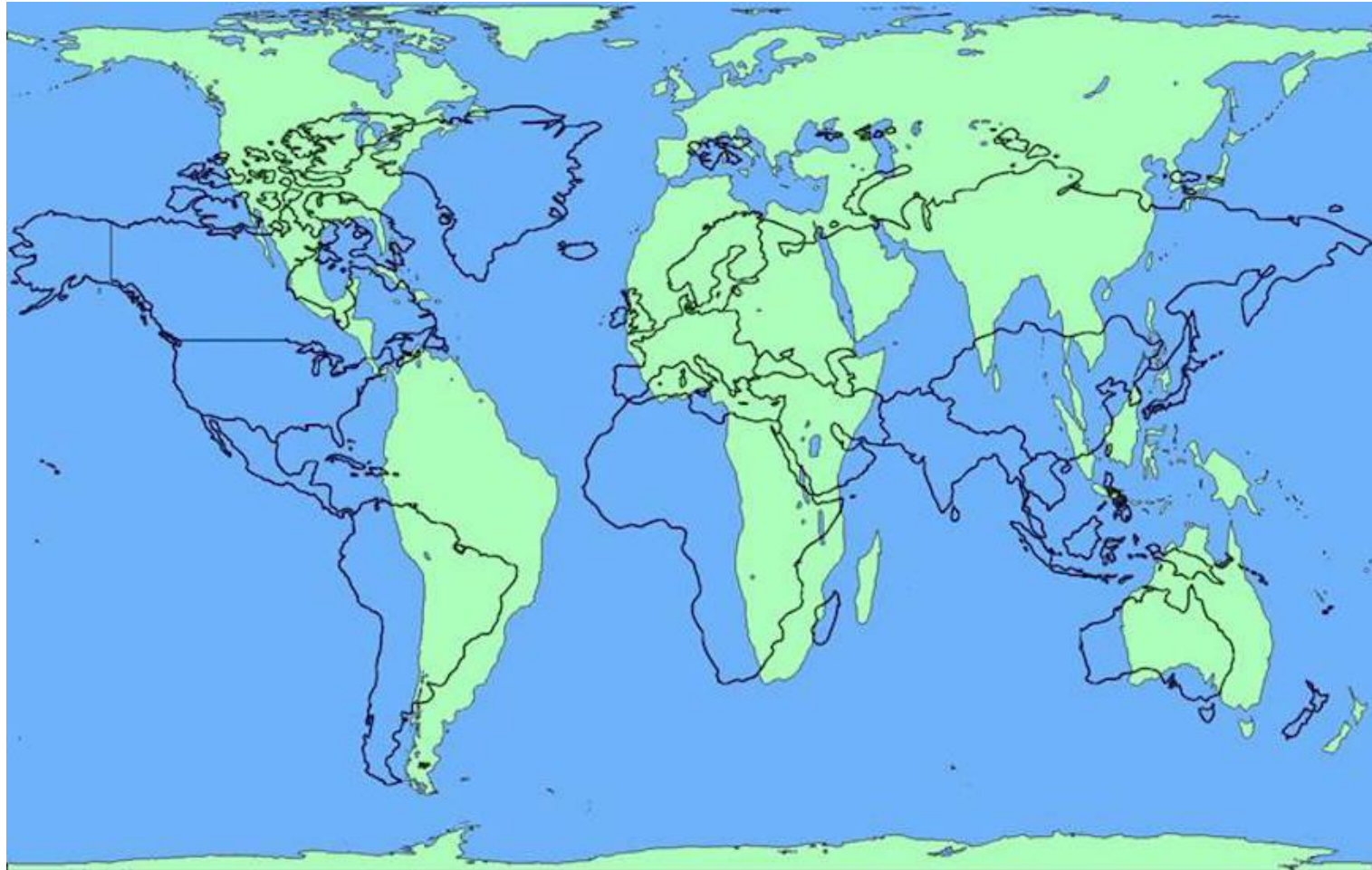
# Designing an Effective Visualization

- Need to take human perception into account (projection/colors)



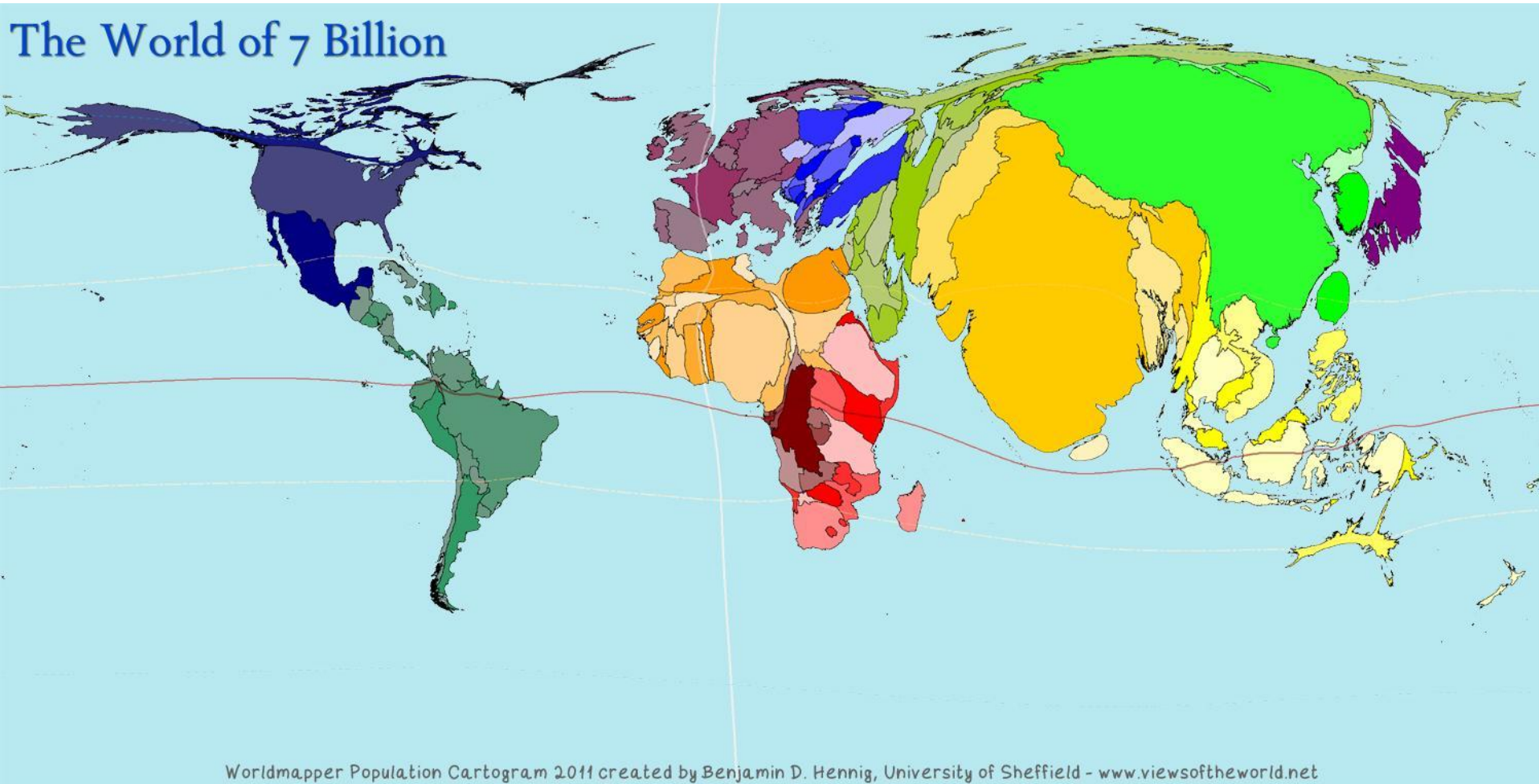
# Designing an Effective Visualization

- Need to take human perception into account (projection)



# Designing an Effective Visualization

- Communicate the right message





# Designing an Effective Visualization

- Consider conflicted entities



# Designing an Effective Visualization

- Consider conflicted entities



# Designing an Effective Visualization



- Human perception is sensitive to:
  - Sizing
  - Colors perception (color choice, clarity, etc)
  - Conflicted entities (names, borders, etc)
  - Values, e.g., population vs population density
  - ...

# Designing an Effective Visualization



- Human perception is sensitive to:
  - Sizing
  - Colors perception (color choice, clarity, etc)
  - Conflicted entities (names, borders, etc)
  - Values, e.g., population vs population density
  - ...
- Visualization confusions might be caused by:
  - Too many colors
  - Inconsistent scales
  - Wrong chart types (e.g., continuous chart on discrete data)
  - ...

# Research on Geo Data Visualization

- A lot of research papers still address big data visualization

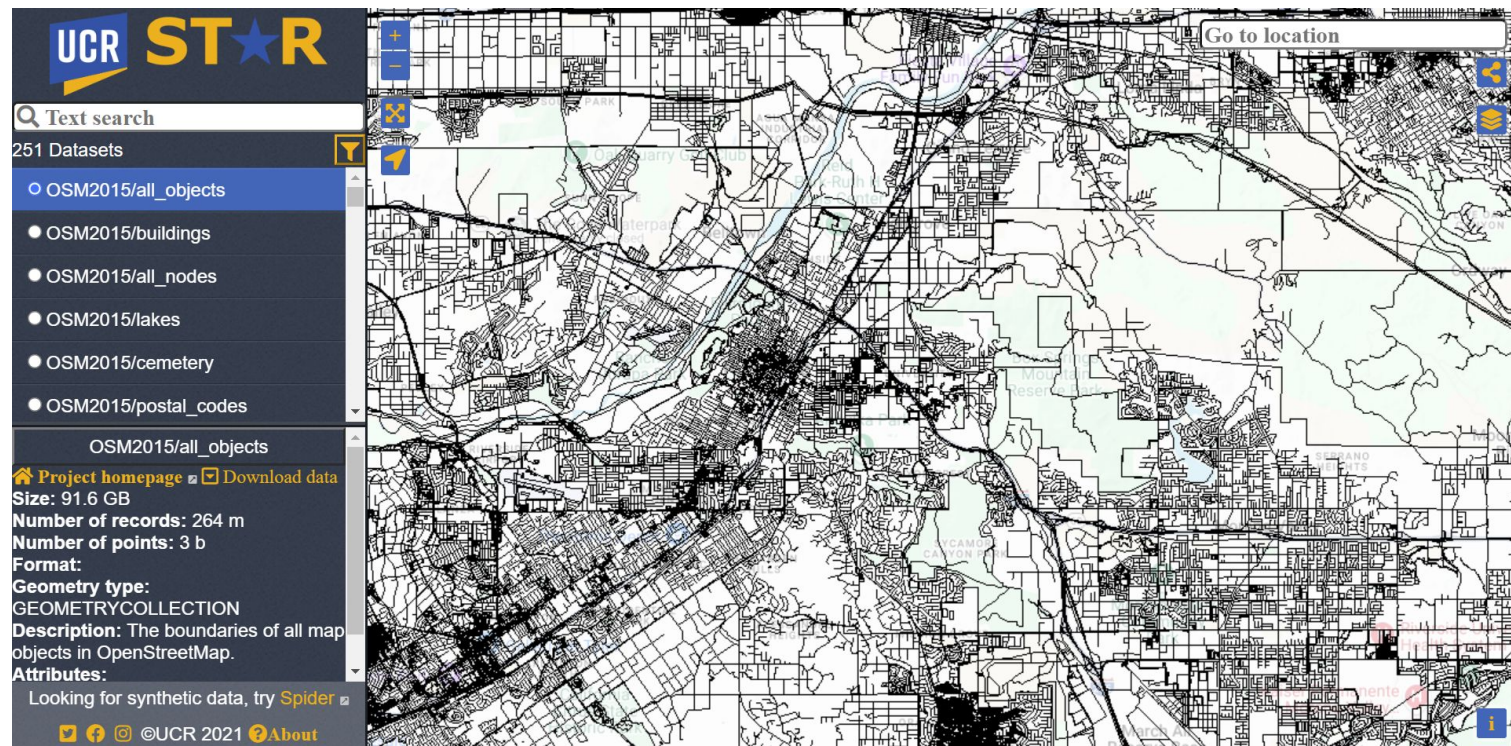
- Example:

AID\*: A Spatial Index for Visual Exploration of Geo-Spatial Data.

By Saheli Ghosh and Ahmed Eldawy

In IEEE TKDE 34(8): 3569-3582 (2022)

- <https://star.cs.ucr.edu/>



The screenshot displays the STAR web application interface. On the left, there is a sidebar with the UCR STAR logo and a text search bar. Below the search bar, a list of 251 datasets is shown, with 'OSM2015/all\_objects' selected. The main area of the interface is a map showing a dense network of lines representing map objects. The sidebar also includes a 'Project homepage' link, a 'Download data' button, and metadata for the selected dataset: 'Size: 91.6 GB', 'Number of records: 264 m', and 'Number of points: 3 b'. The 'Description' field states: 'The boundaries of all map objects in OpenStreetMap.' The 'Attributes' field is currently empty. At the bottom of the sidebar, there are social media icons and the text 'Looking for synthetic data, try Spider'.

# Challenges

## 1. Number of Datasets

Filter by location  Clear

141,168 datasets found

**Collection 1 meter Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection** [1,054 recent views](#)

*U.S. Geological Survey, Department of the Interior* — This is a tiled collection of the 3D Elevation Program (3DEP) and is one meter resolution. The 3DEP data holdings serve as the elevation layer of The National Map...

WAF IMG Esri REST WMS HTML HTML Regional

**Zip Codes - ZipCodes** [910 recent views](#)

*NSGIC State / GIS Inventory* — This dataset represents an ongoing effort to approximate the geographic extents of 5 digit zip codes. The dataset was produced using a combination of methods and is...

XML ZIP ZIP HTML State

**U.S. Hourly Precipitation Data** [855 recent views](#)

*National Oceanic and Atmospheric Administration, Department of Commerce* — Hourly Precipitation Data (HPD) is digital data set DSI-3240, archived at the National Climatic Data Center (NCDC). The primary source of data for this file is...

HTML HTML HTML HTML Esri REST XML 11 more in dataset Regional

**Food Environment Atlas** [367 recent views](#)

*Department of Agriculture* — Food environment factors—such as store/restaurant proximity, food prices, food and nutrition assistance programs, and community characteristics—interact to...

HTML ZIP HTML Regional

**TIGER/Line Shapefile, 2017, nation, U.S., Current State and Equivalent National** [552 recent views](#)

*US Census Bureau, Department of Commerce* — The TIGER/Line shapefiles and related database files (.dbf) are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master...

ZIP HTML WMS Esri REST

**Pittsburgh Wards Map** [448 recent views](#)

*Allegheny County / City of Pittsburgh / Western PA Regional Data Center* — Allows users to look up City of Pittsburgh Wards

## 2. Big Data

eBird

Project homepage

Download

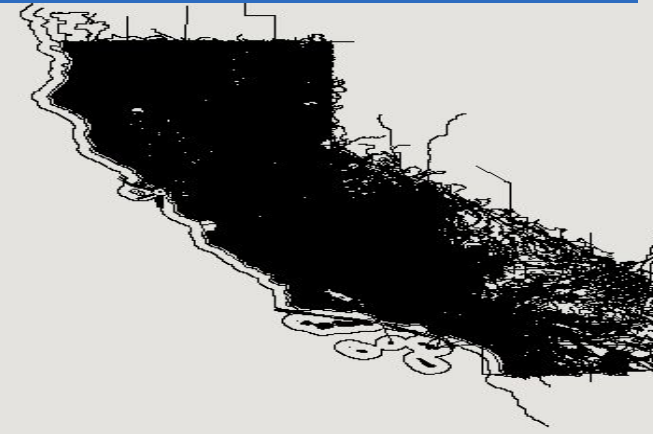
Size: 211.2 GB

Number of records: 566

Format: CSV

Geometry type

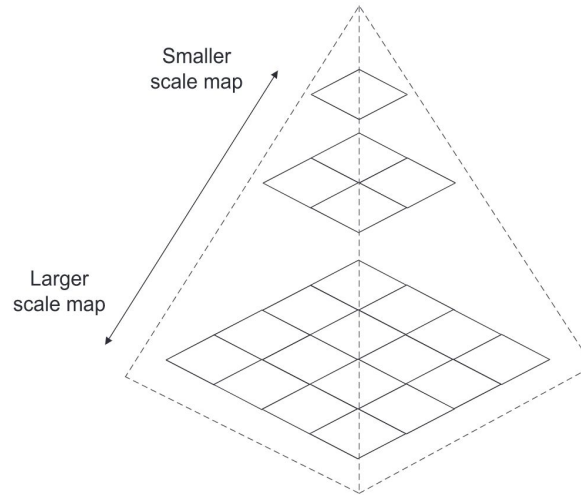
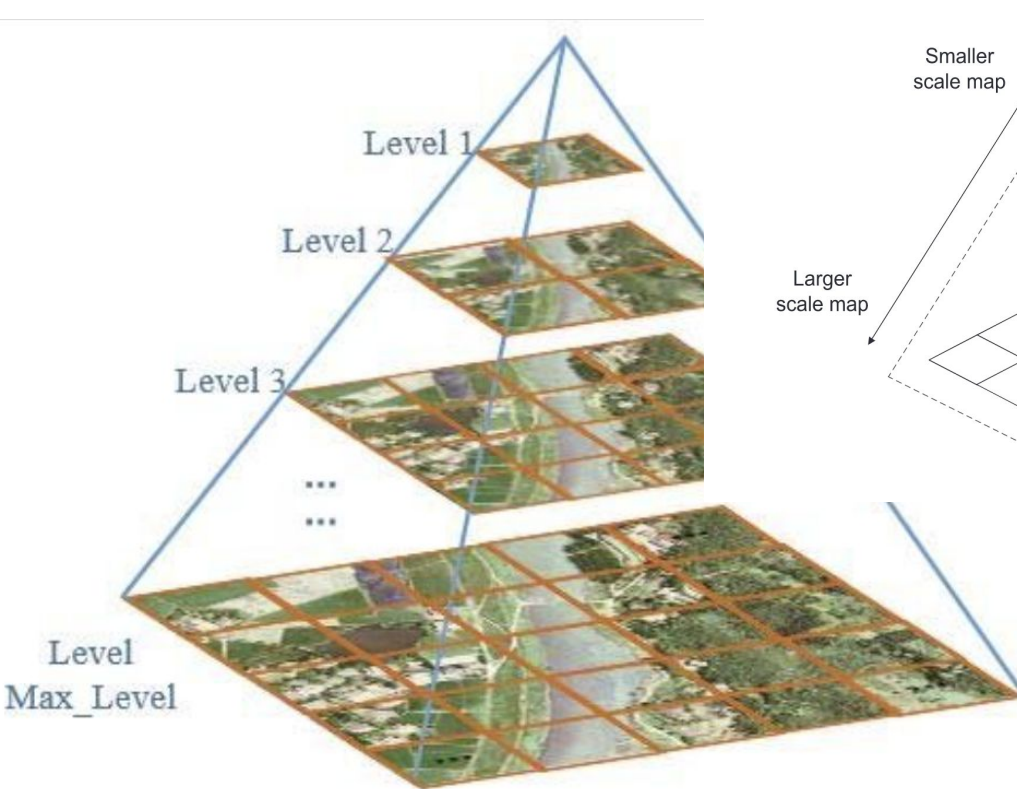
## 3. Interactivity



## 4. Cost Effectiveness



# Multi-level Visualization



Tile Structure:

Level 0: 1 tile +

Level 1: 4 tiles +

Level 2: 16 tiles + .....

Level 10: 1,048,576 +

Level 20: More than billion tiles

Traditional quad-tree - HadoopViz,  
GeoSparkViz, Google Maps, Open  
Street Maps

Not scalable for multiple datasets, deeper zoom levels or larger data

# Multi-level Visualization using Vector Data

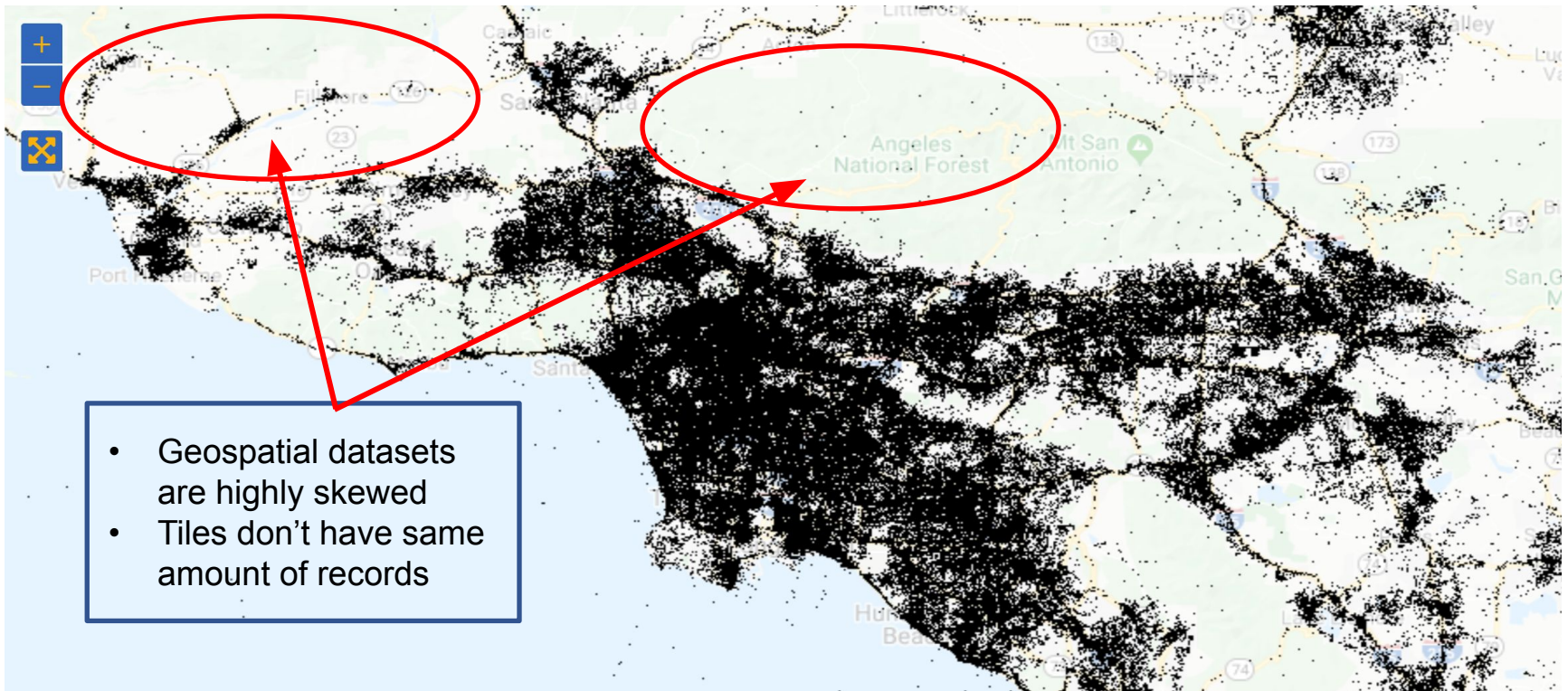


- Vector data: Raw data with attributes like locations (lat, lon), shape, etc
- Images are generated on-the-fly every time a tile is requested by the user

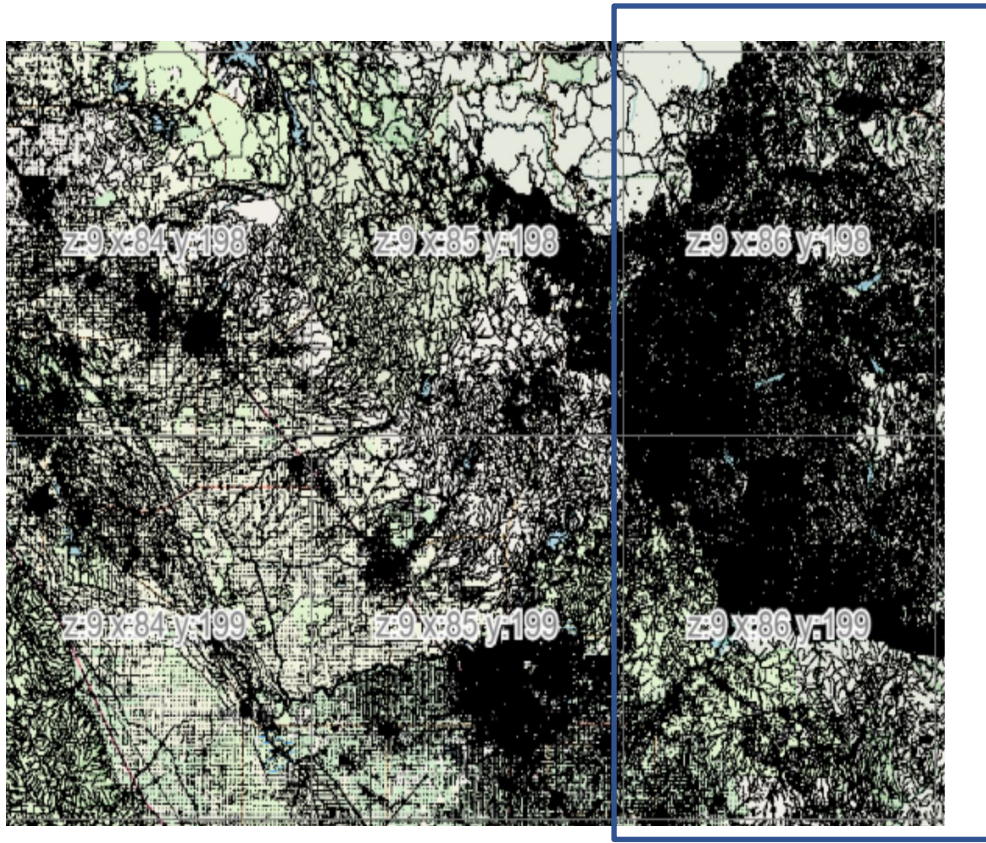
If a tile contains too much records, it hinders the interactivity.  
Cannot scale for big data.



# Typical Geospatial Datasets



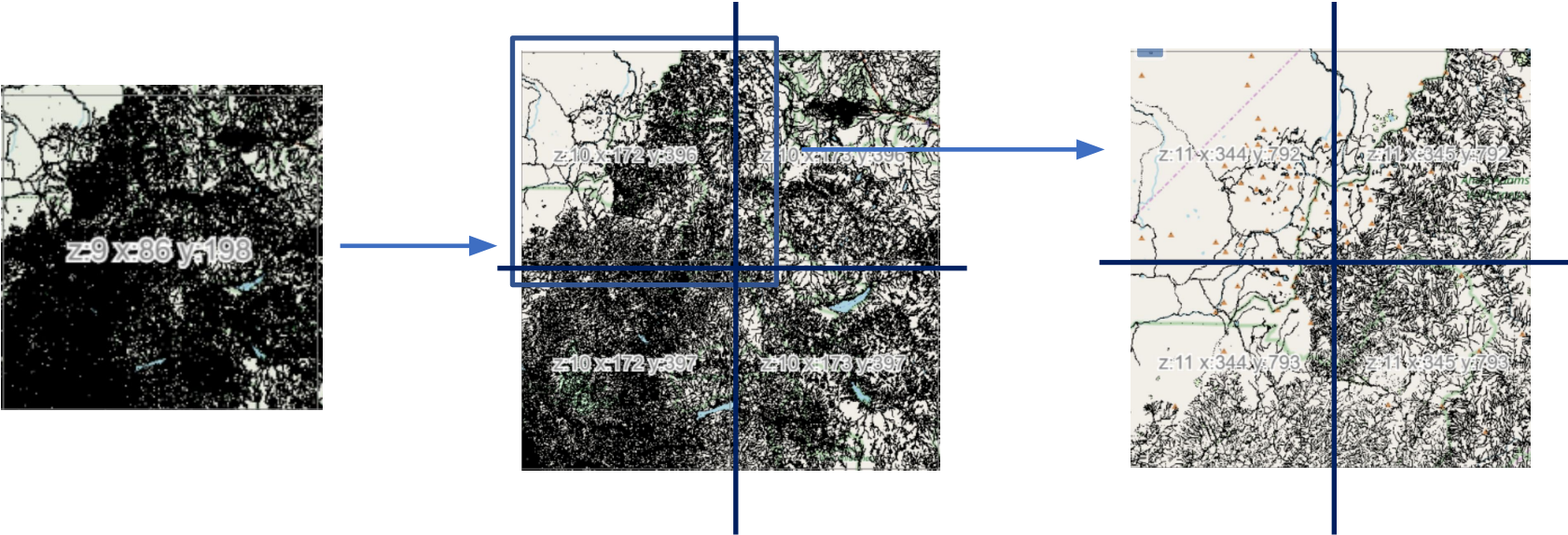
# Varying tile density across the area



Tiles has more dense records compared to the remaining four



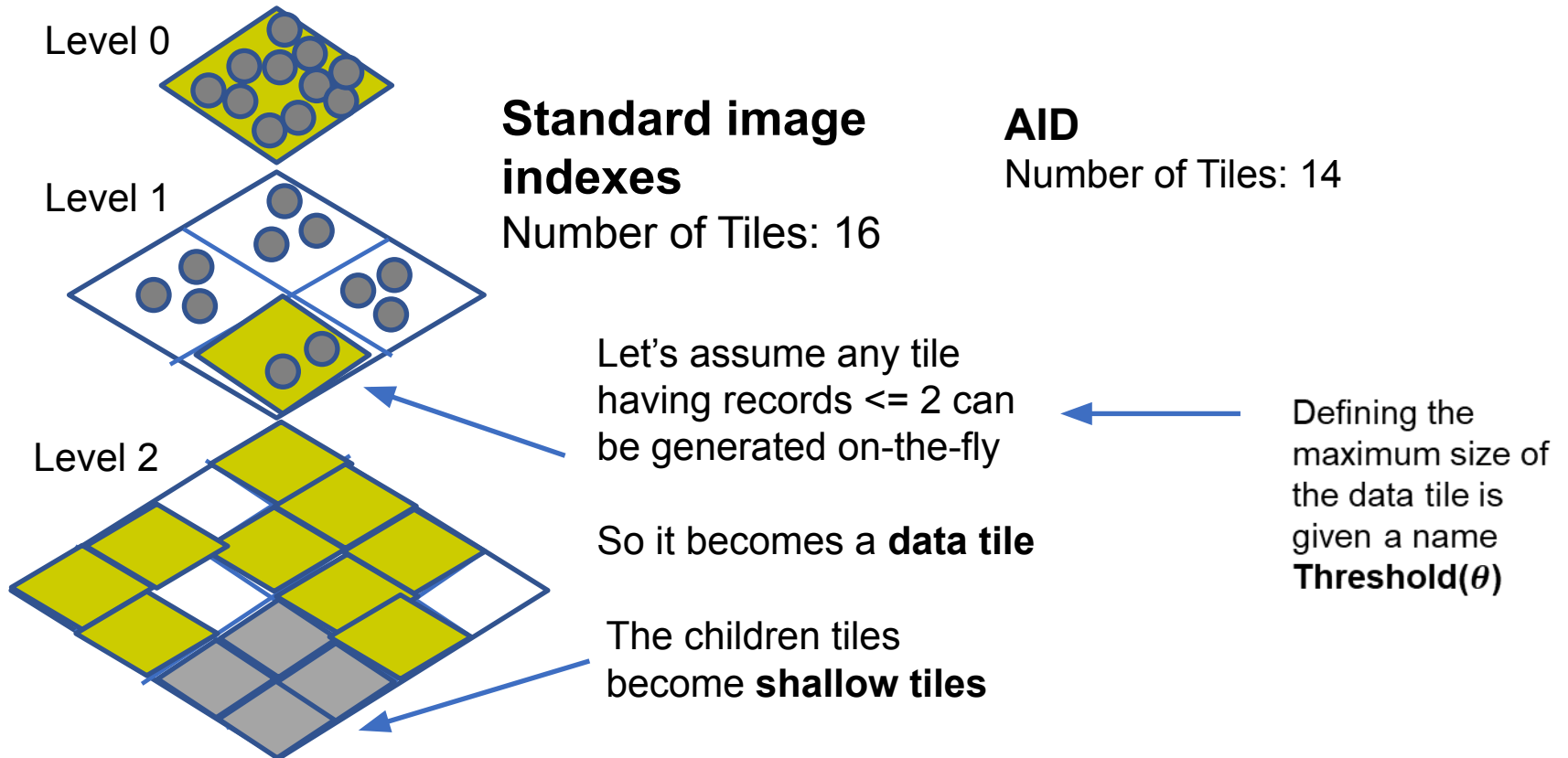
# Tile density across zoom levels



# AID/AID\* Index

- Classifying tiles according to their size (or amount of data they hold) to build an adaptive index
- Pregenerating the heavy, dense, record-filled tiles
- Generate the tiles with fewer records on-the-fly

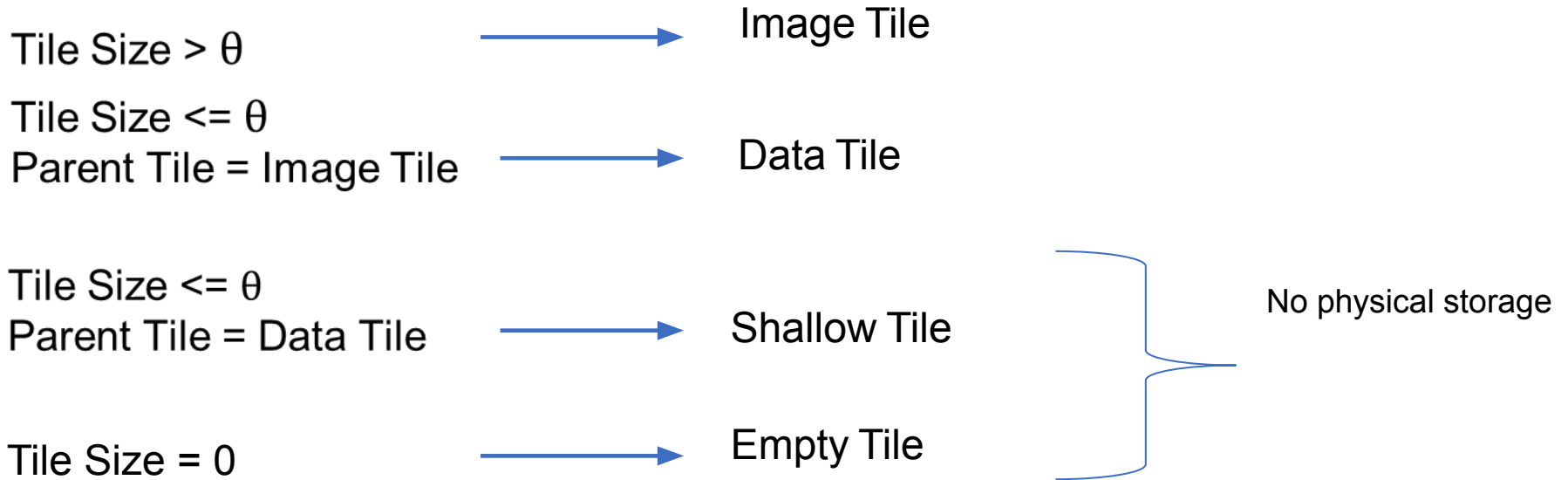
# Example of AID

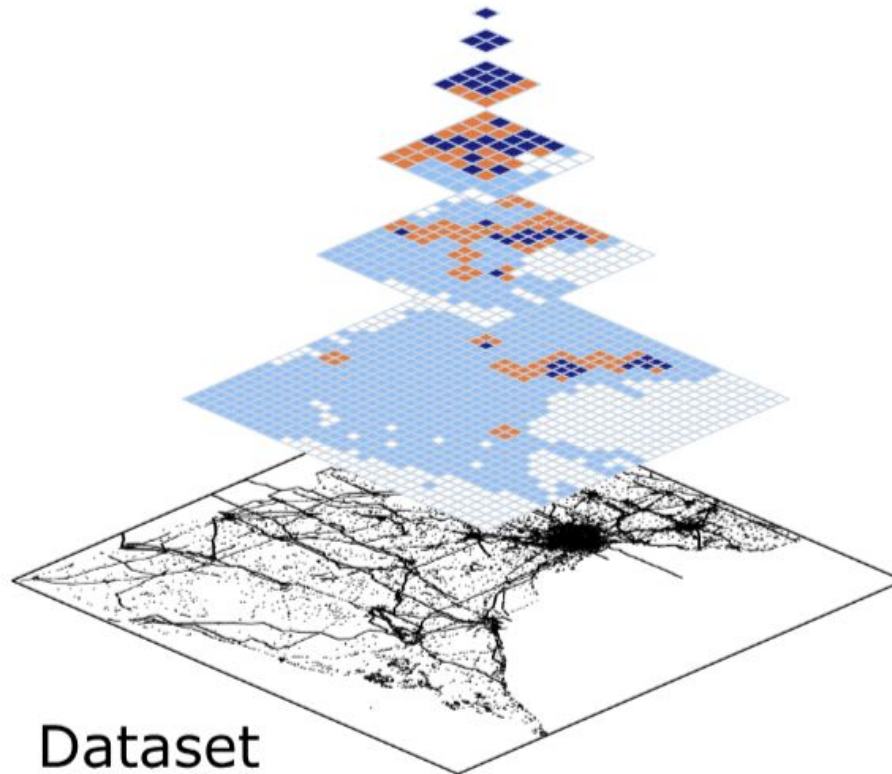


# Threshold for tile classification



A parameter based on the size of each tile

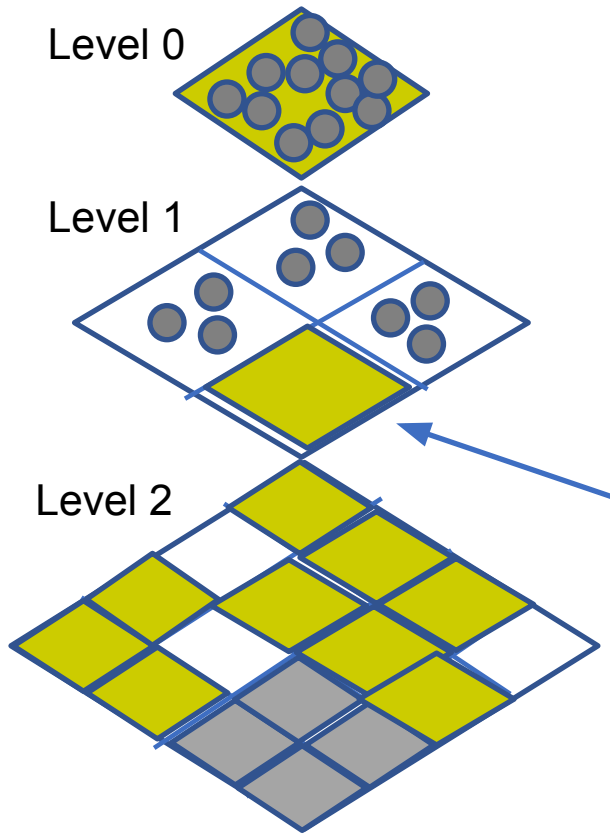




## Dataset

- ◆ Image tiles (43 tiles)
- ◆ Data tiles (92 tiles)
- ◆ Shallow tiles (1230 tiles)

# AID\*



**AID**

Number of Tiles: 14

**AID\***

Number of Tiles: 4

AID\* does not create or store data tiles



# AID\*

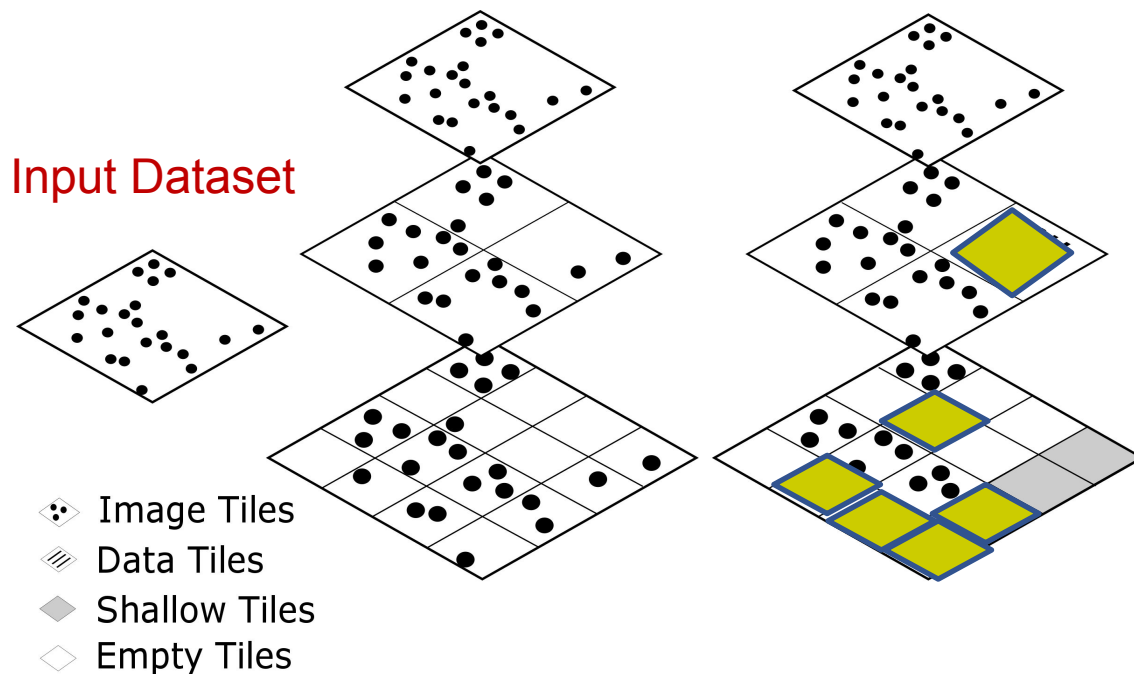
- Previously indexed dataset (R\* index)
- A pyramidal quad-tree having pregenerated image tiles
- Not materializing data tiles
- Indexed overhead reduced to 0.01%

# Index Construction

- Data summarization
  - Compute a histogram to summarize the data to calculate the size of each tile
- Tile classification
  - Classify the tiles as image, data, shallow or empty using the histogram
- Tile creation
  - Based on the tile classification image tiles are pregenerated as .png files, data tiles are created and stored as .csv files(AID) and shallow and empty tiles have no physical storage. AID\* does not store data tiles as well

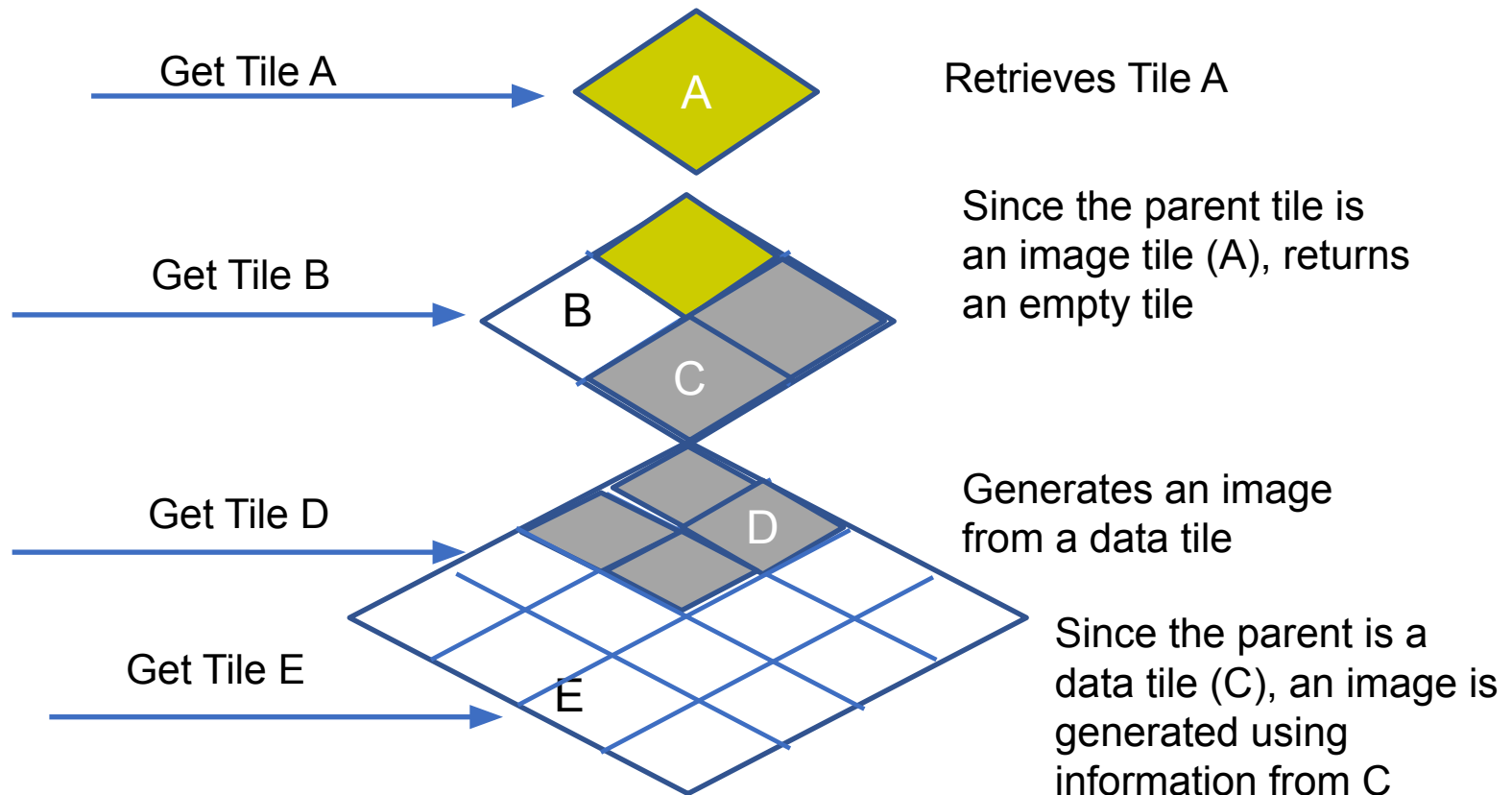
# Quadtree vs. AID vs. AID\*

Traditional quad-tree image index **AID\***

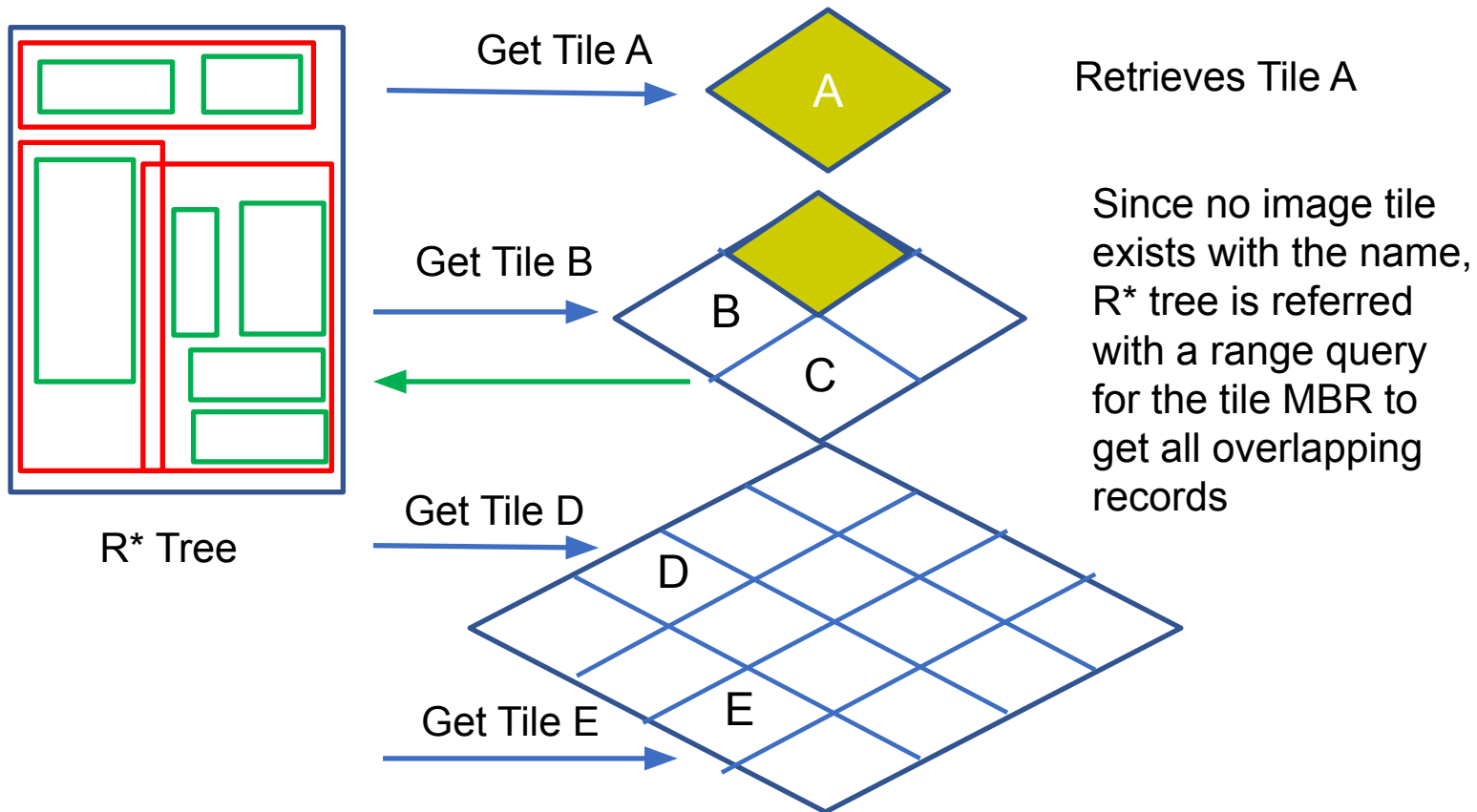


Traditional Index: 15  
tiles  
AID: 13 tiles  
AID\*: 8 tiles

# Visualization Query for AID (Single-Machine)

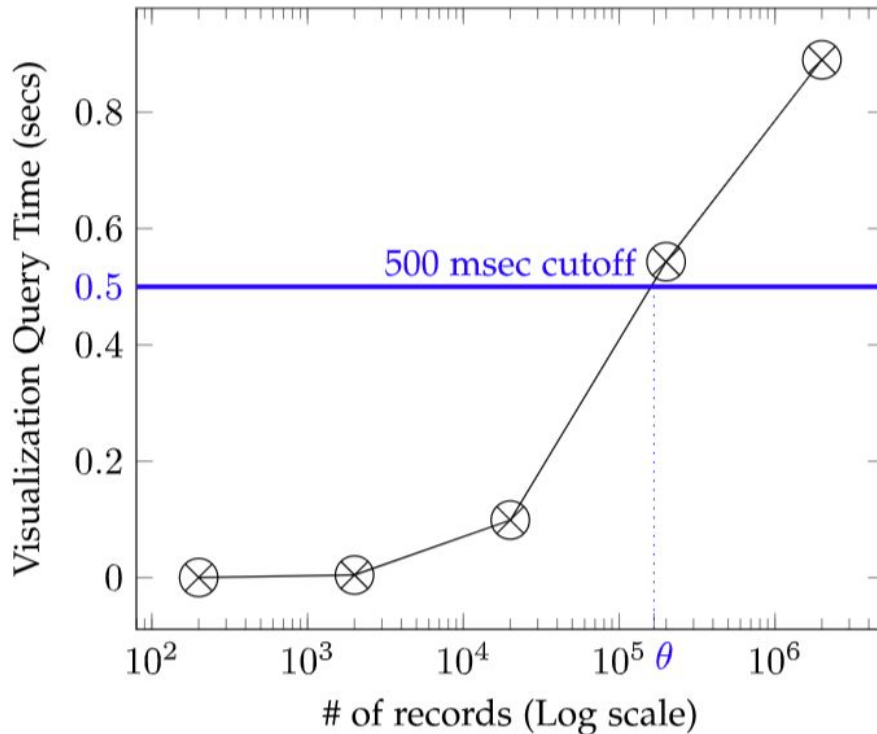


# Visualization Query for AID\* (Single-Machine)



# Tuning Interactivity

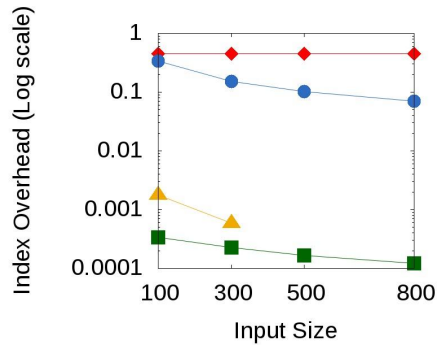
- Through tuning the threshold  $\theta$



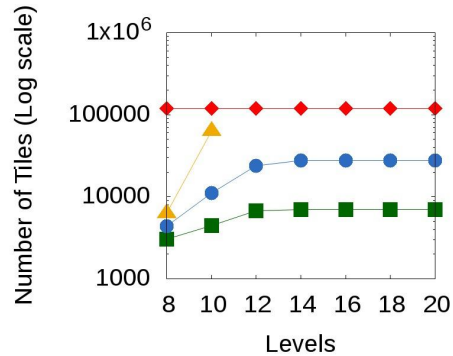
- Size of  $\theta$  is the biggest size a data tile can be
- Too big  $\theta \Rightarrow$  low interactivity
- Too small  $\theta \Rightarrow$  exponentially growing image index

Visualization query time with increasing number of records in a single tile

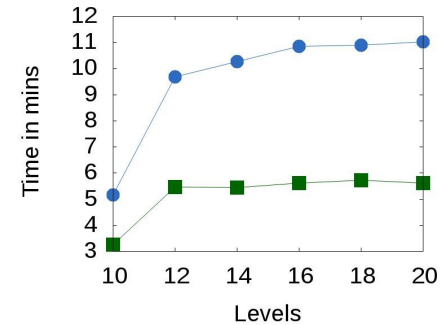
# Experiments



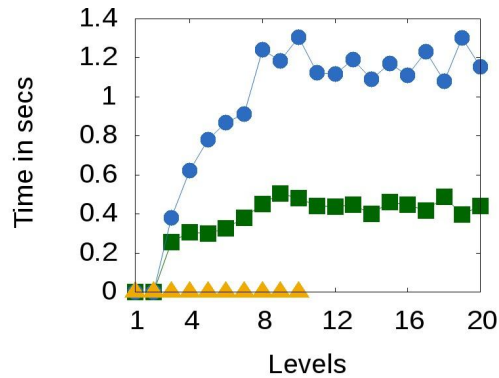
**Input Size Vs Index Overhead**



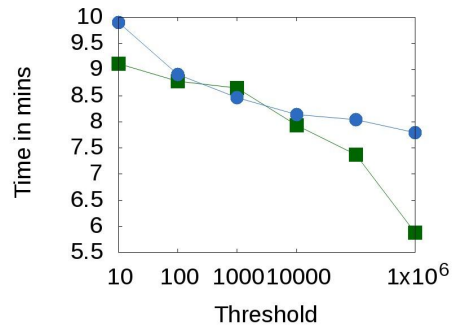
**Levels Vs Number of Tiles**



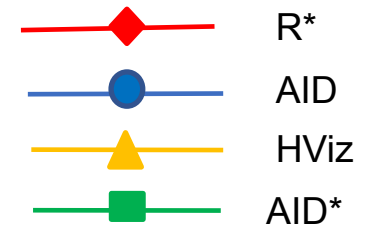
**Levels Vs Const Time**



**Levels Vs Vis Query Time**



**Threshold Vs Const Time**



# Credits

- Prof. Luc Anselin's lecture
  - <https://www.youtube.com/watch?v=KJFSSET0Diw>
- Prof. Ahmed Eldawy and Dr. Saheli Ghosh work
- Dr. Ning Guo work