UCR

Spatial Computing

UCR

UNIVERSITY OF CALIFORNIA
UC RIVERSIDE

Geovisualization

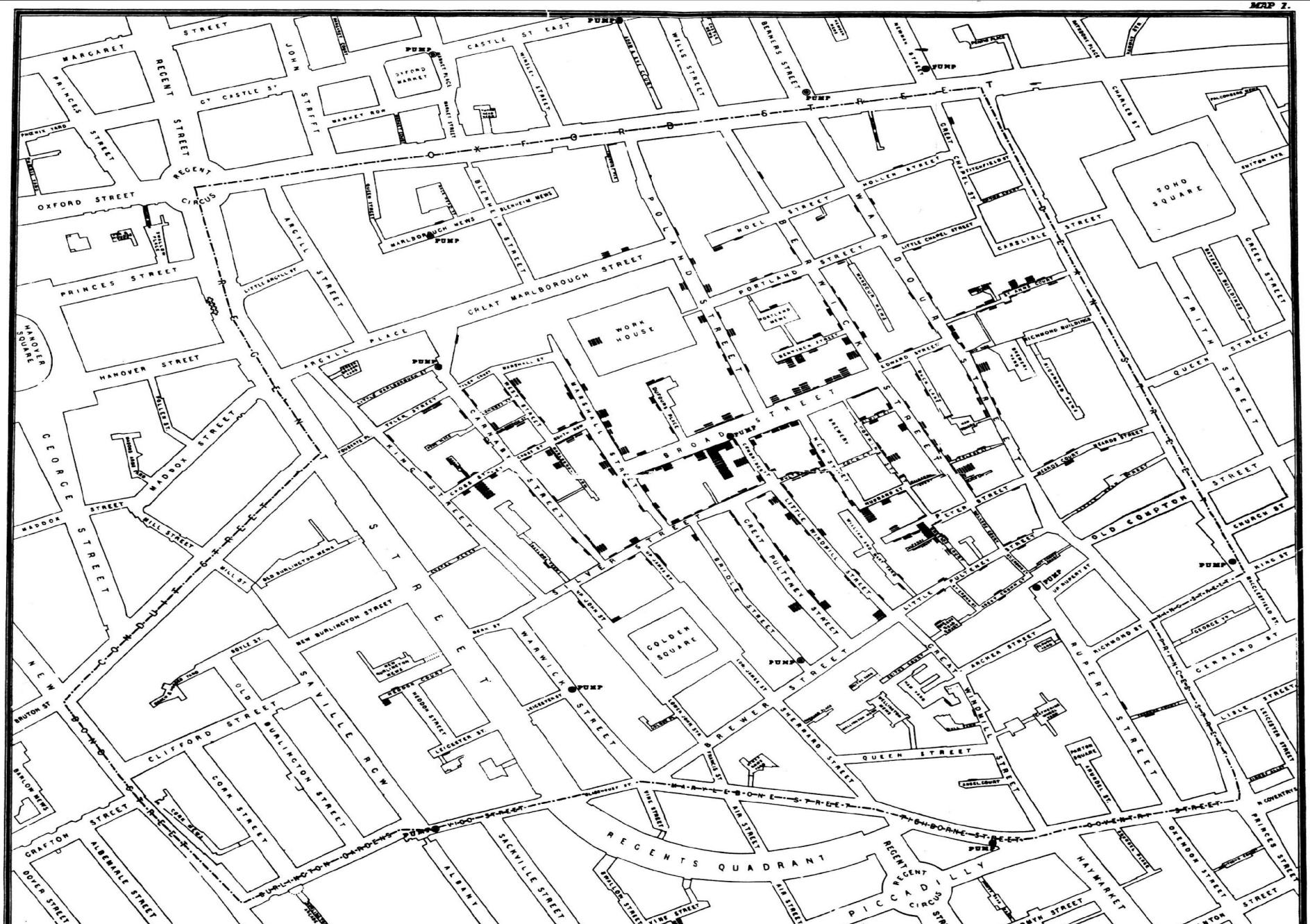
Amr Magdy

Computer Science and Engineering Department
Center for Geospatial Sciences
University of California, Riverside
www.cs.ucr.edu/~amr/

Visual Perception

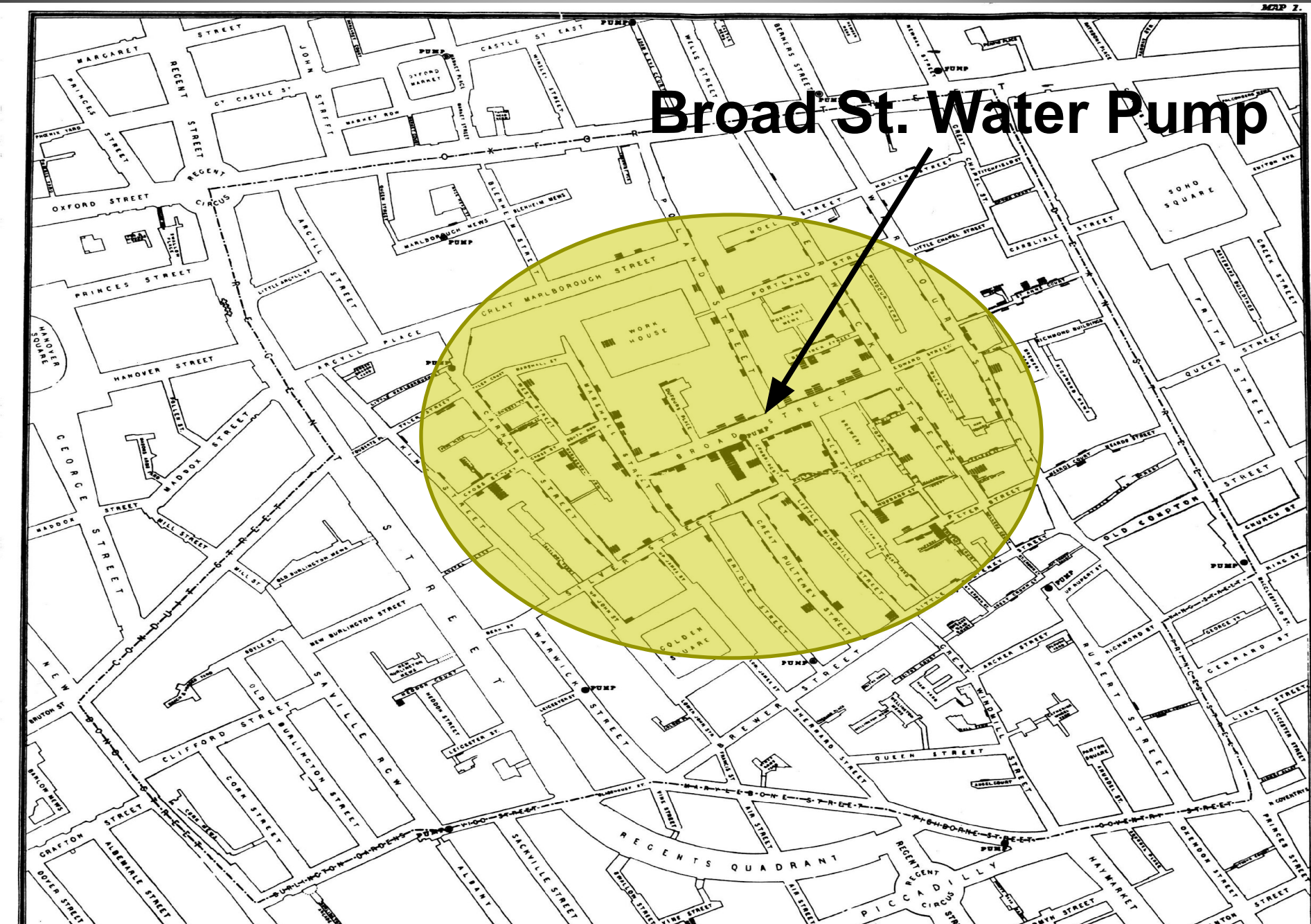
- Learning Styles & Personality Types: Visual, Auditory, Kinesthetic

Cholera cases in the London epidemic of 1854



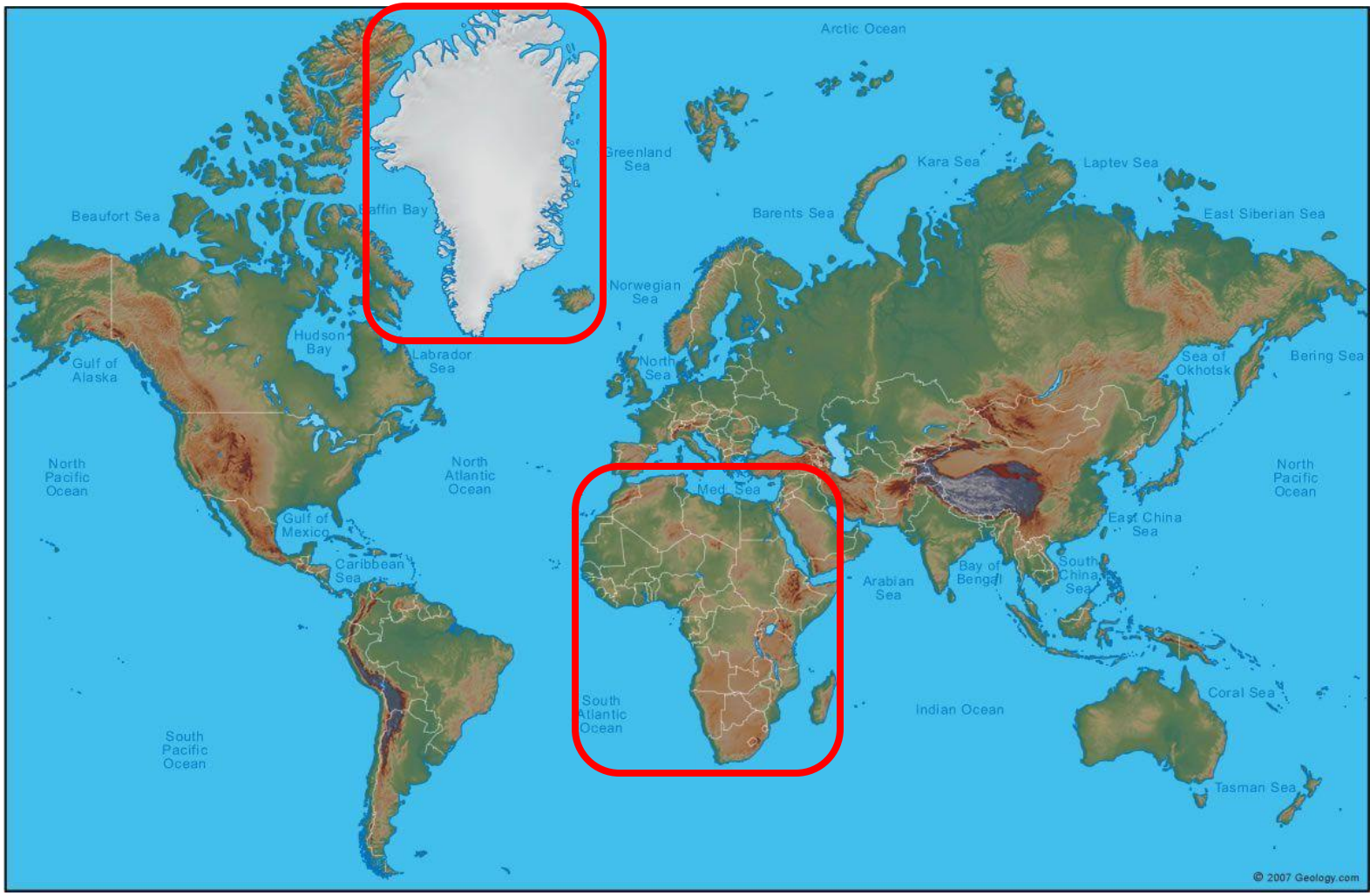
Cholera cases in the London epidemic of 1854

Cholera cases in the London epidemic of 1854



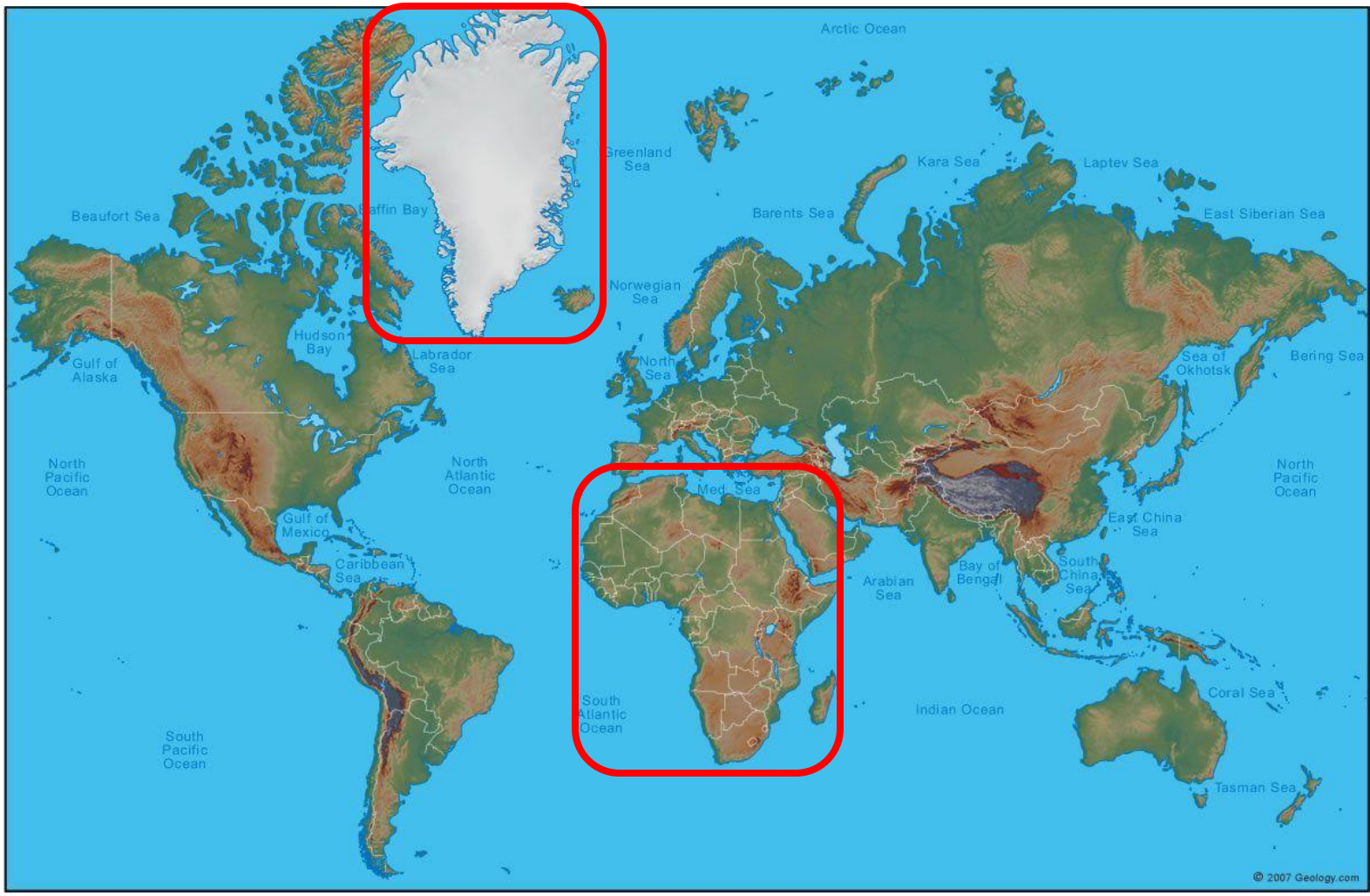
Geo-Visualization

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



Geo-Visualization

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

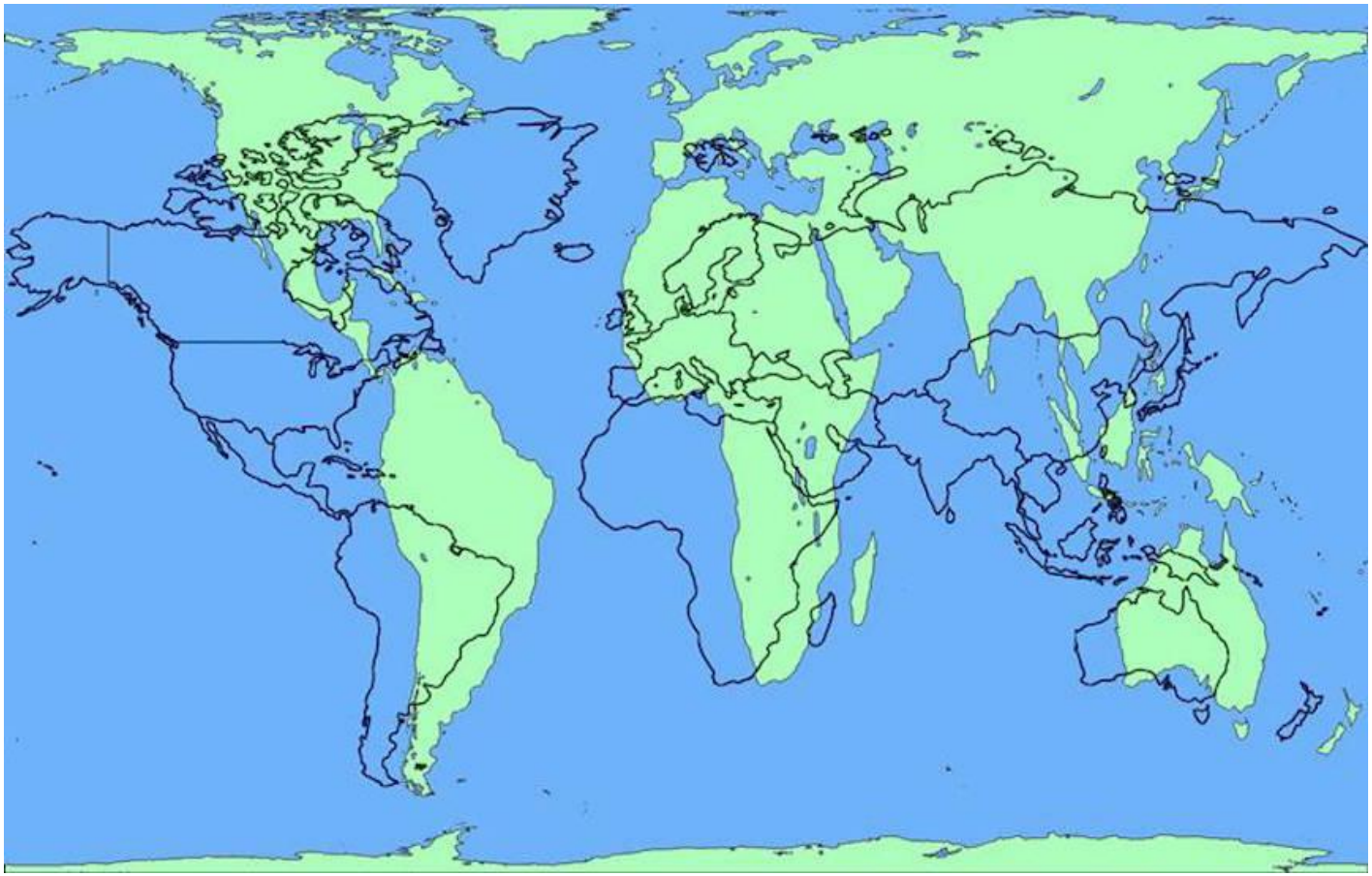


Map Orientation and Projections

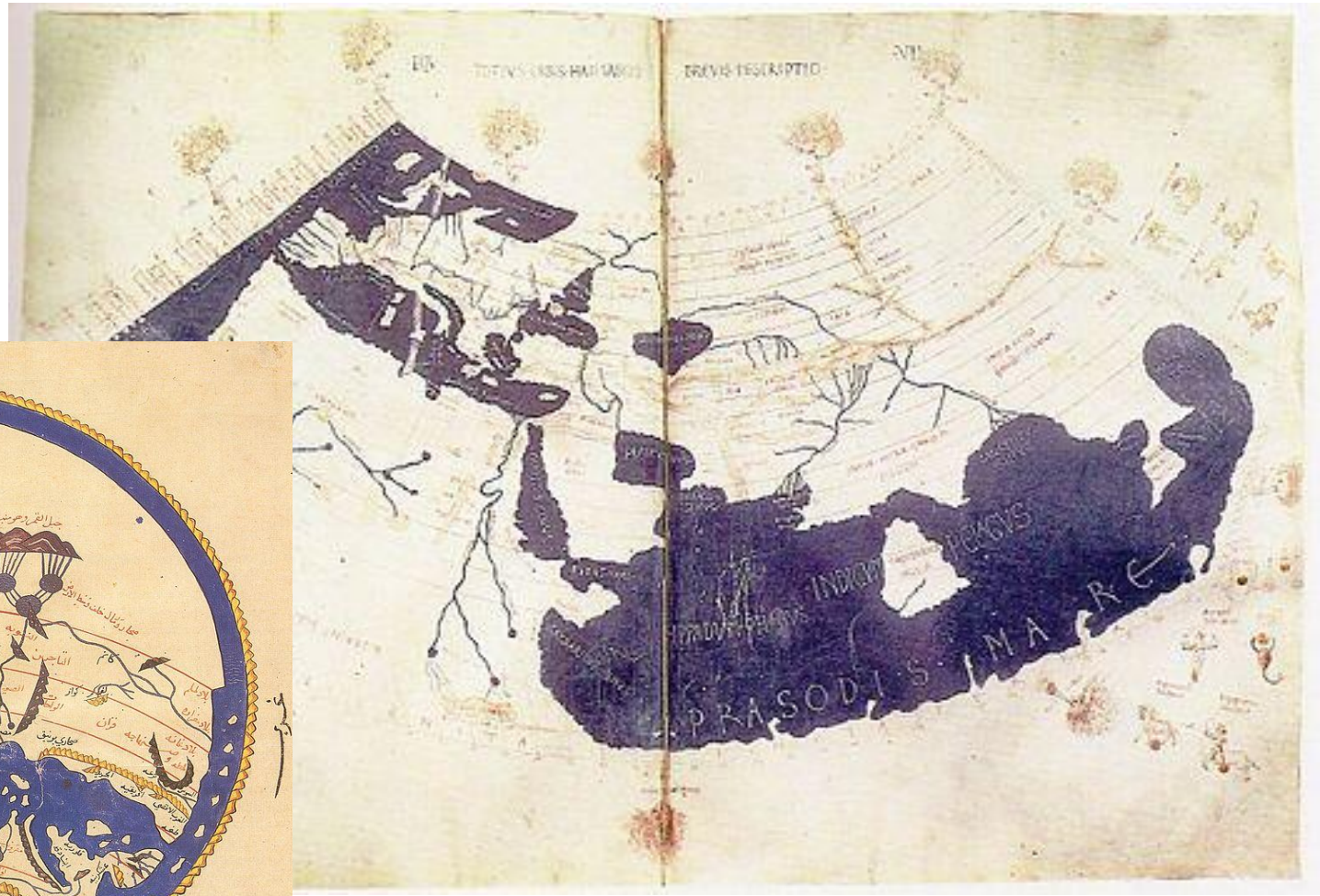
- Mapping a 3D globe on a flat 2D plane
  - Why all world maps are wrong?
    - https://www.youtube.com/watch?v=kIID5FDi2JQ

Map Orientation and Projections

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



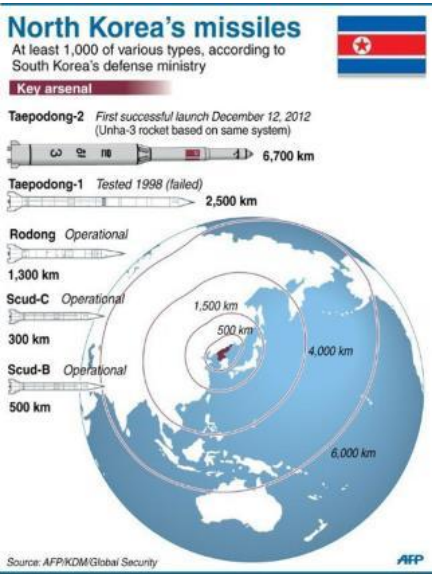
Map Orientation and Projections



Map Orientation and Projections

Map Orientation and Projections

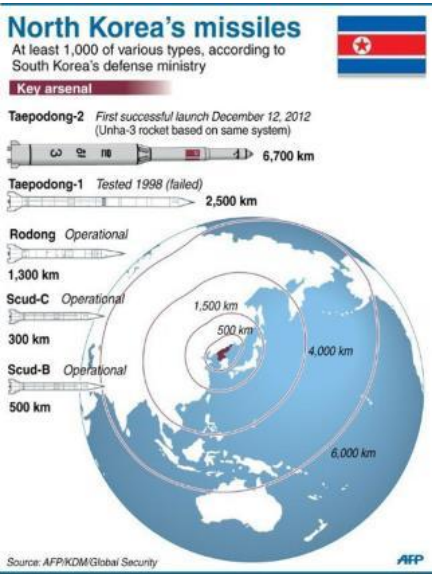
The Economist



Map Orientation and Projections

The Economist

Original



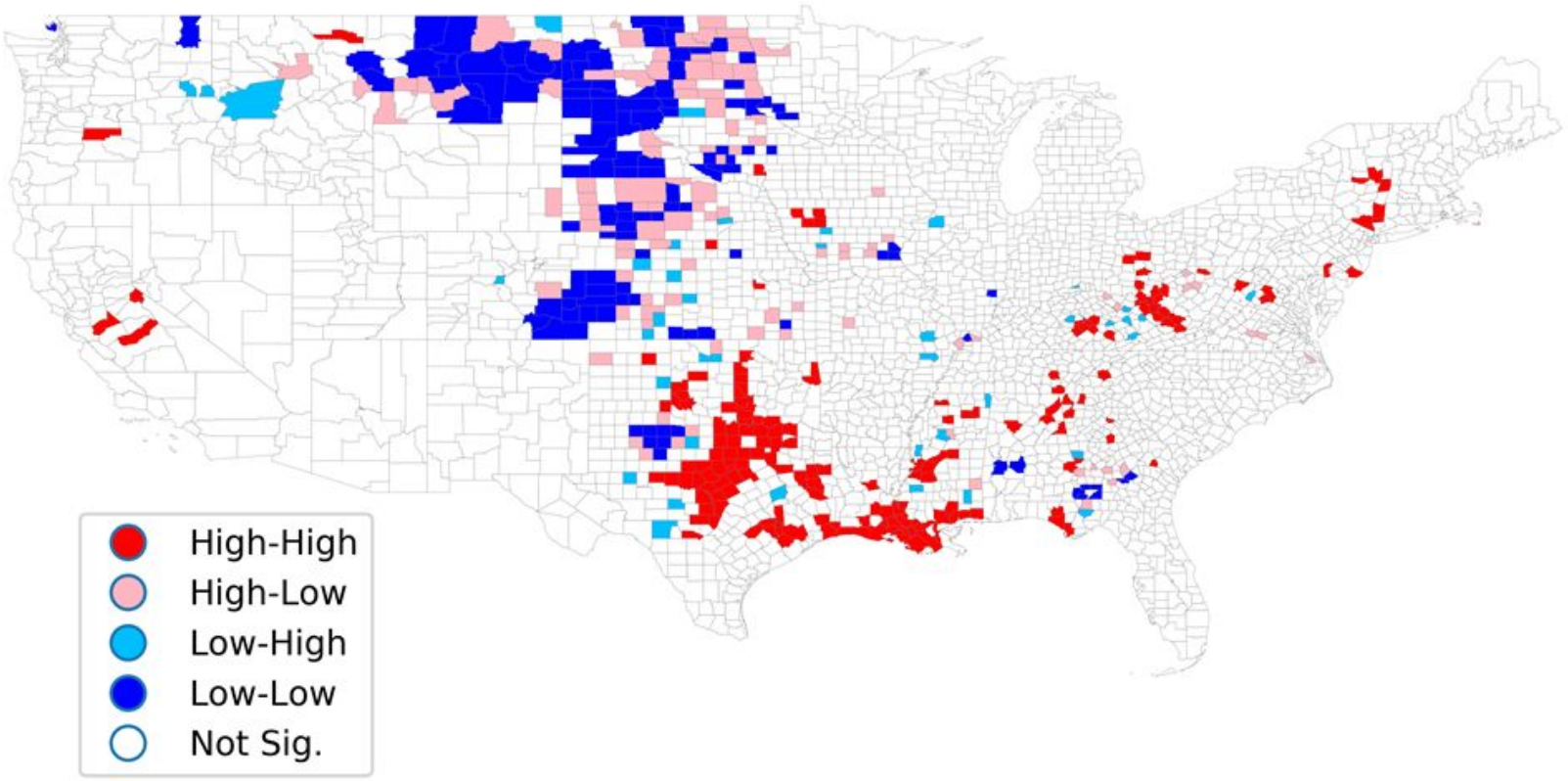
Correction

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

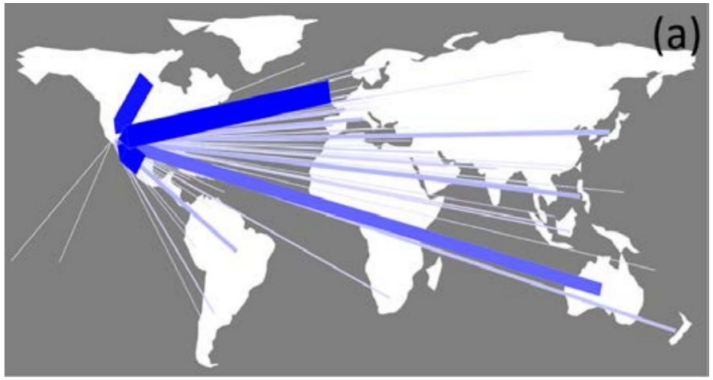
- Mapping
  - With all map applications throughout history
- Decision making
  - E.g., disease outbreaks, crimes, etc
- Real-time monitoring
  - E.g., traffic, security, etc
- Scientific analysis
  - E.g., climate change, vegetation analysis, etc
- …

Geo-visualization Element

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

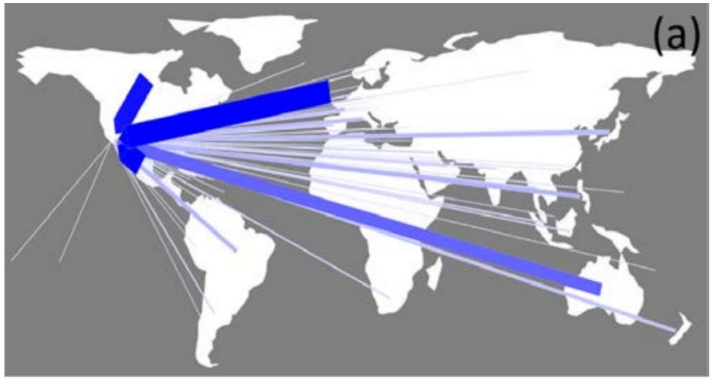
Geo-visualization Element

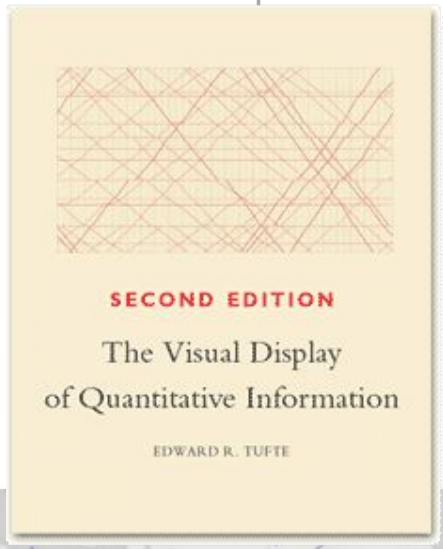
- Three elements
  - Data: what to visualize?
  - Location: where to put data?
  - Visualization scheme: how to visualize?

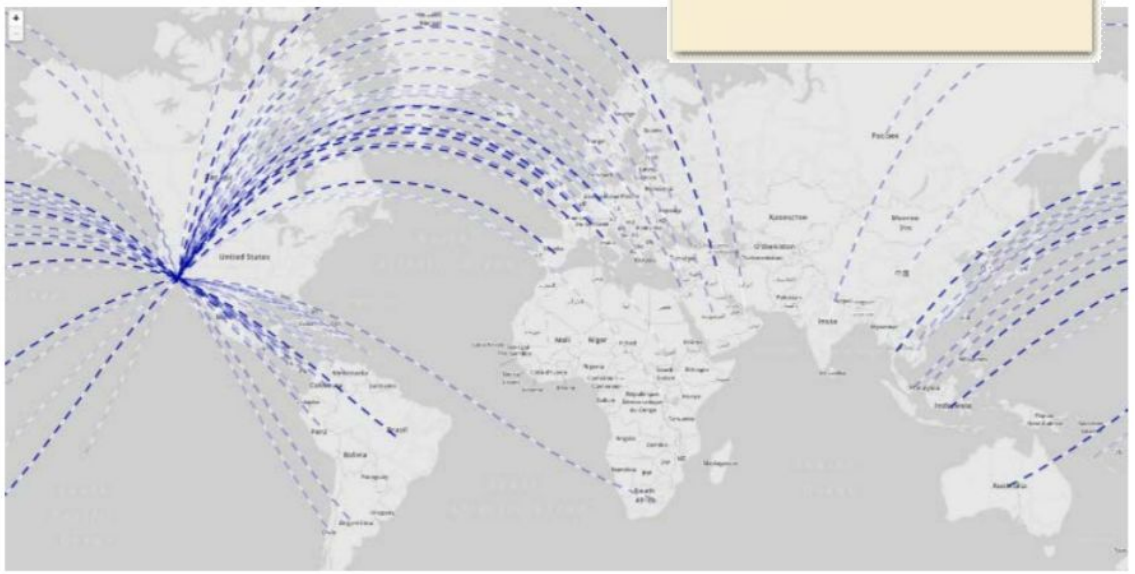


Geo-visualization Element

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

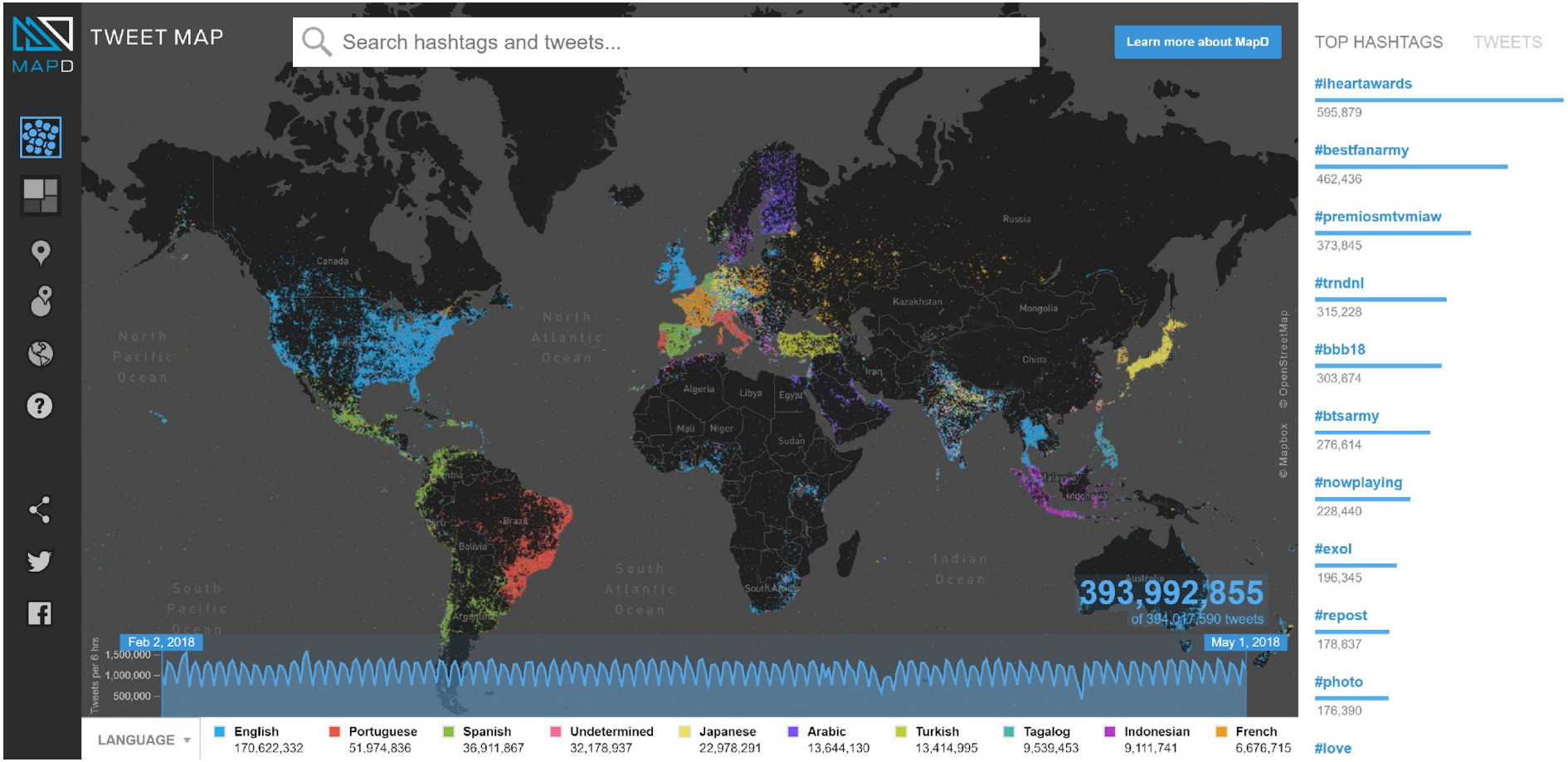






Interactive Maps

- MapD interactive demos
  - Tweet map: https://www.mapd.com/demos/tweetmap/



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)

Visualization in Virtual Reality

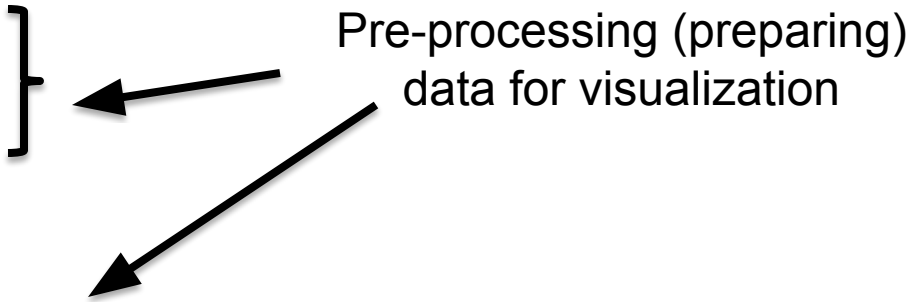
- https://www.youtube.com/watch?v=u76ww3NJFgE

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

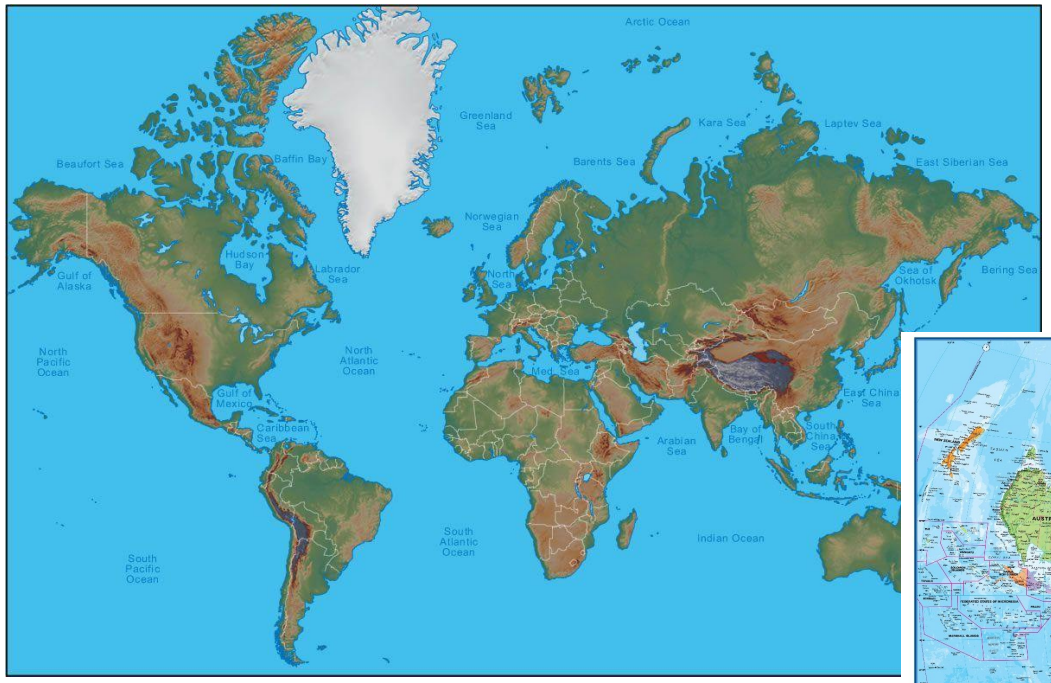
- New challenges come with big volume data
  - How to put data on the map?
  - How to aggregate large data?
  - How to process large data?



- High velocity
  - High velocity data visualization exploits pre-materialization
  - Still active research is on-going

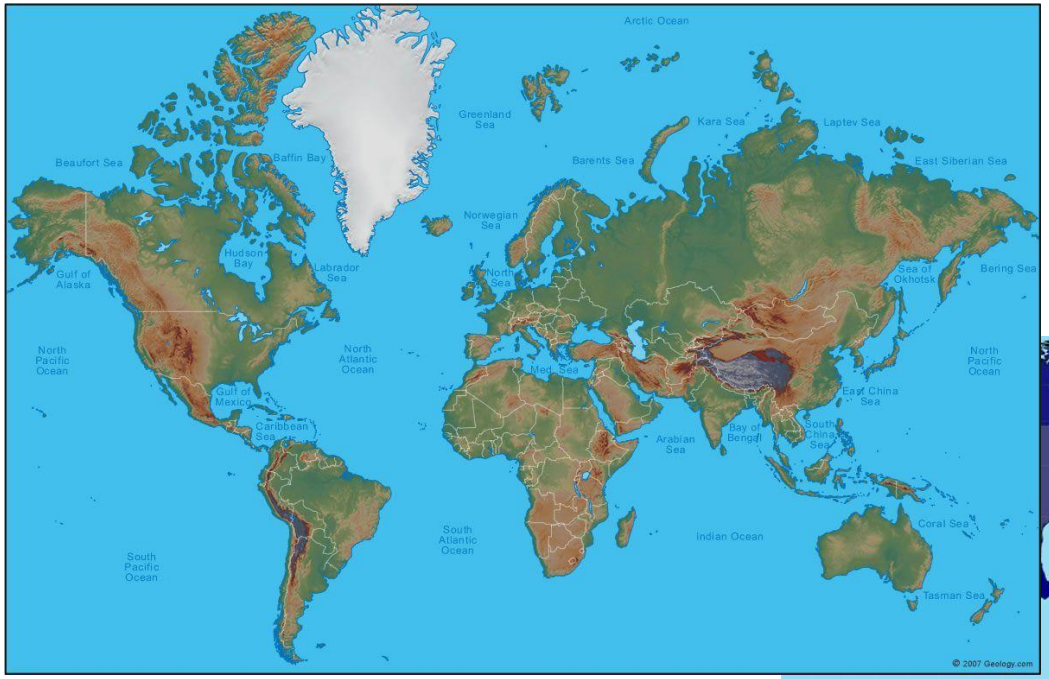
Designing an Effective Visualization

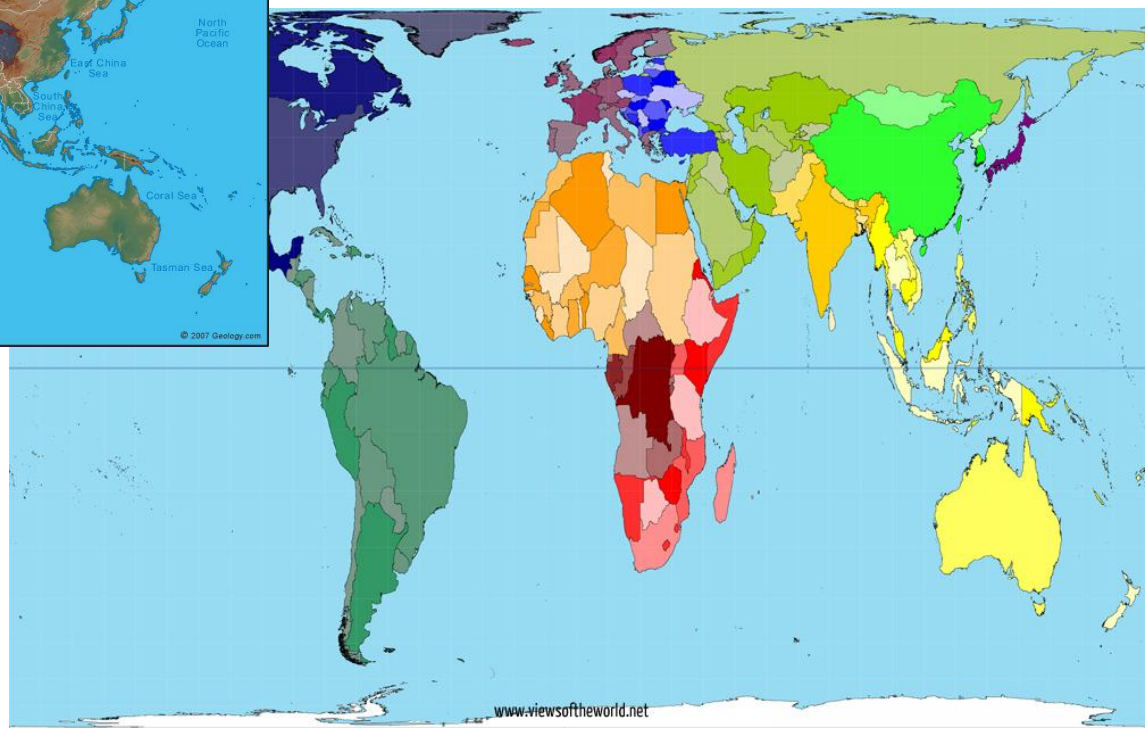
- Need to take human perception into account (orientation)



Designing an Effective Visualization

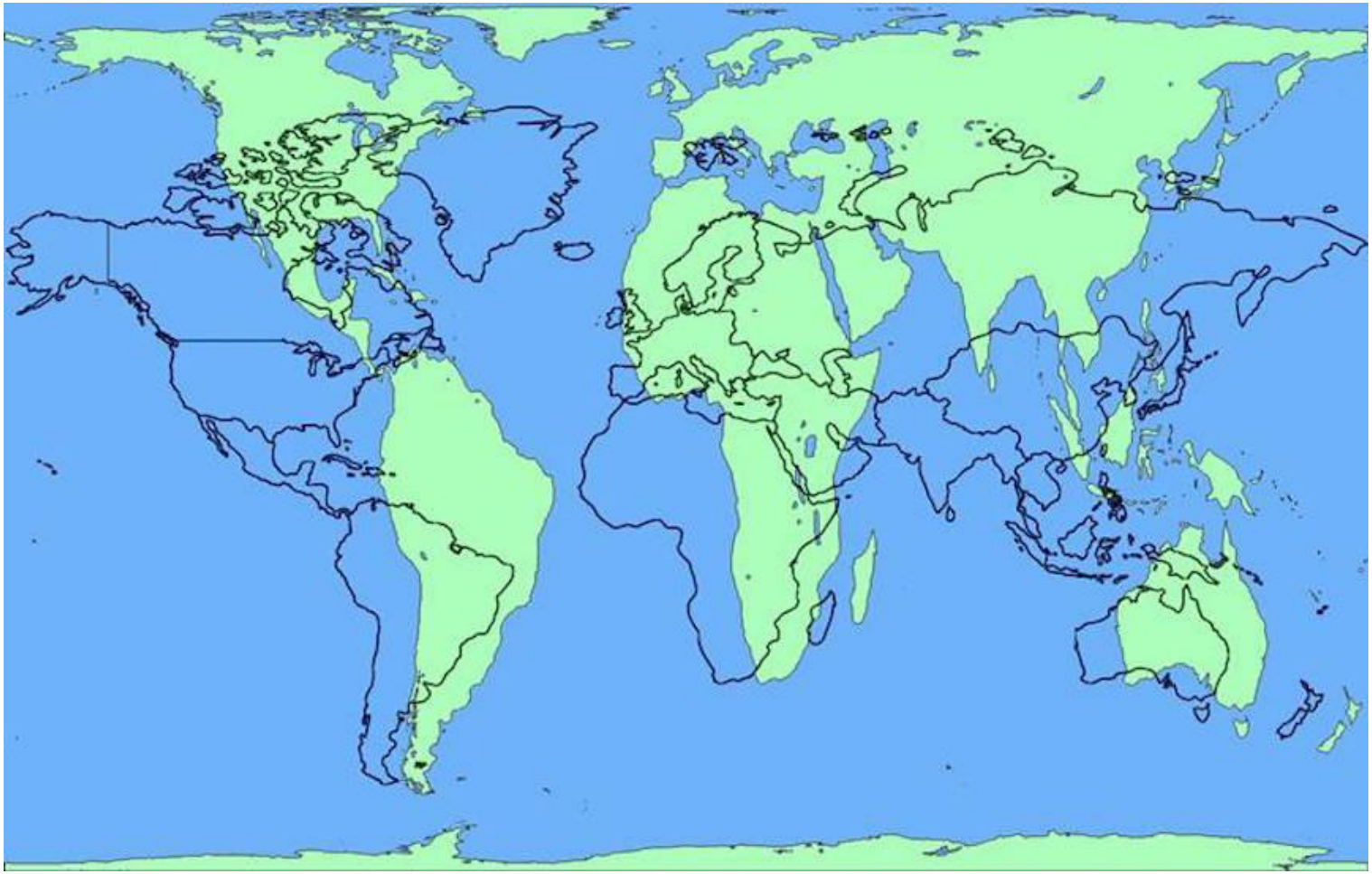
- Need to take human perception into account (projection/colors)





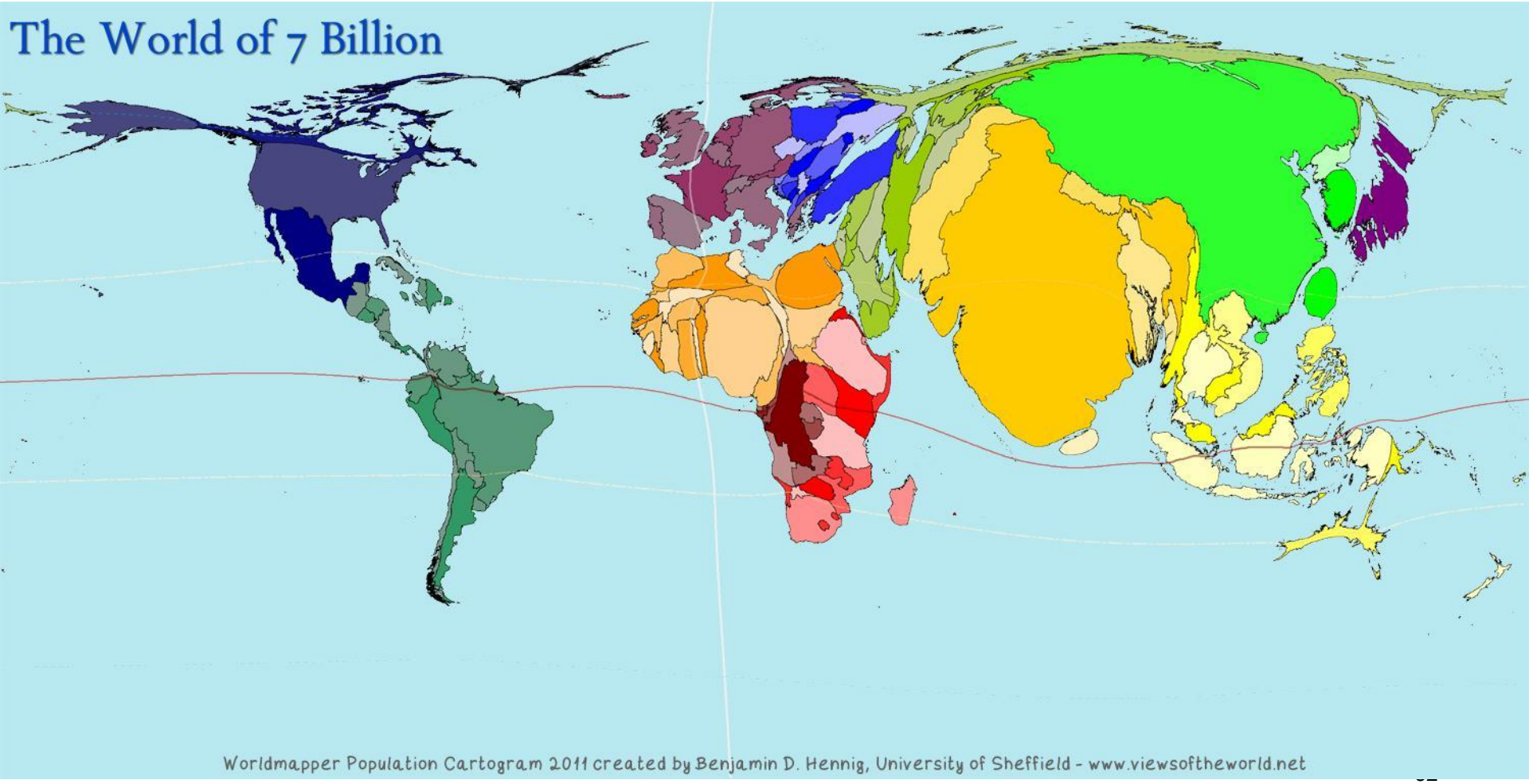
Designing an Effective Visualization

- 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

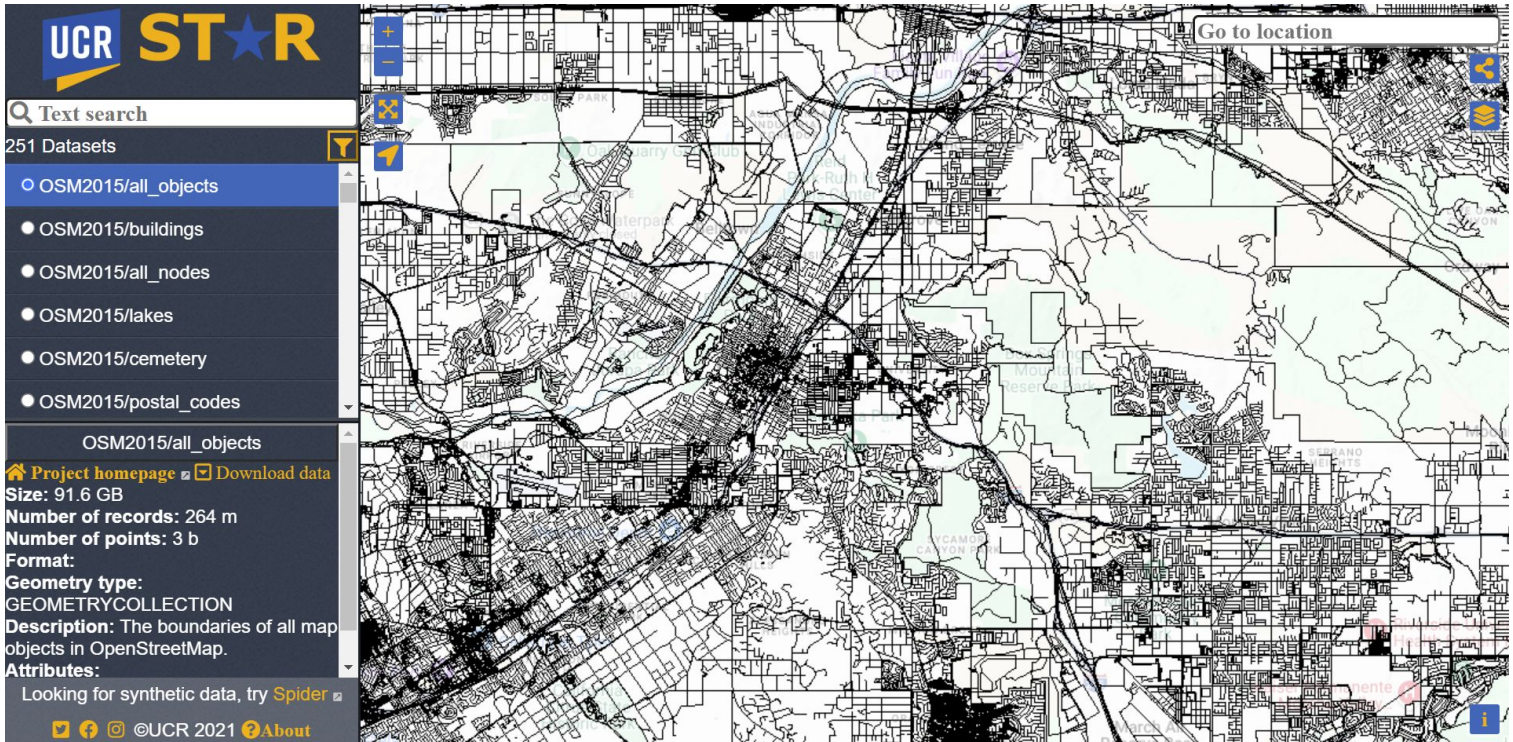
- Human perception is sensitive to:
  - Sizing
  - Colors perception (color choice, clarity, etc)
  - Conflicted entities (names, borders, etc)
  - Values, e.g., population vs population density
  - …
- Visualization confusions might be caused by:
  - Too many colors
  - Inconsistent scales
  - Wrong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

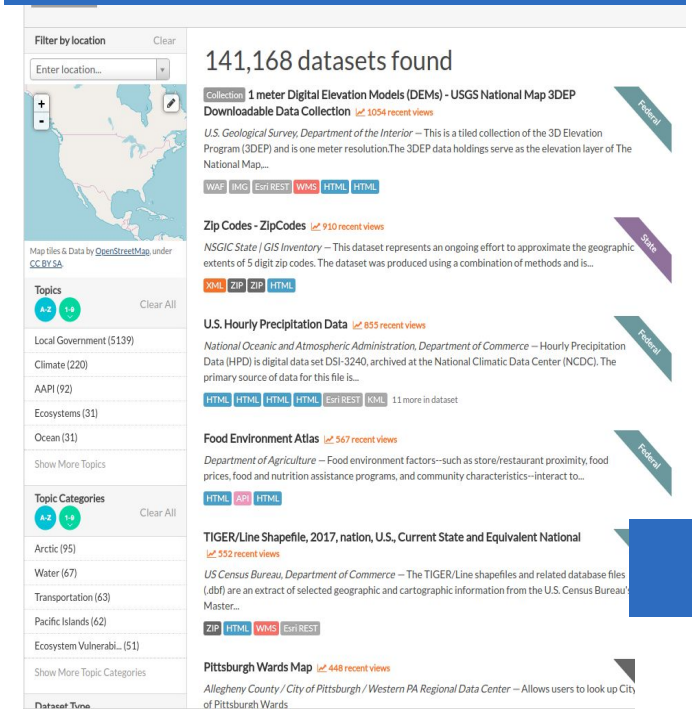
AID*: A Spatial Index for Visual Exploration of Geo-Spatial Data.
By Saheli Ghosh and Ahmed Eldawy
In IEEE TKDE 34(8): 3569-3582 (2022)

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

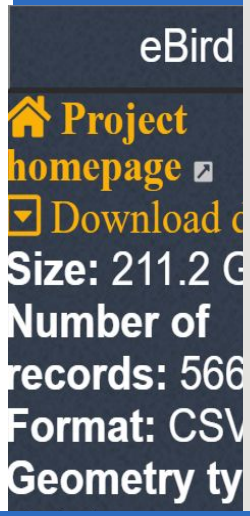


Challenges

1. Number of Datasets



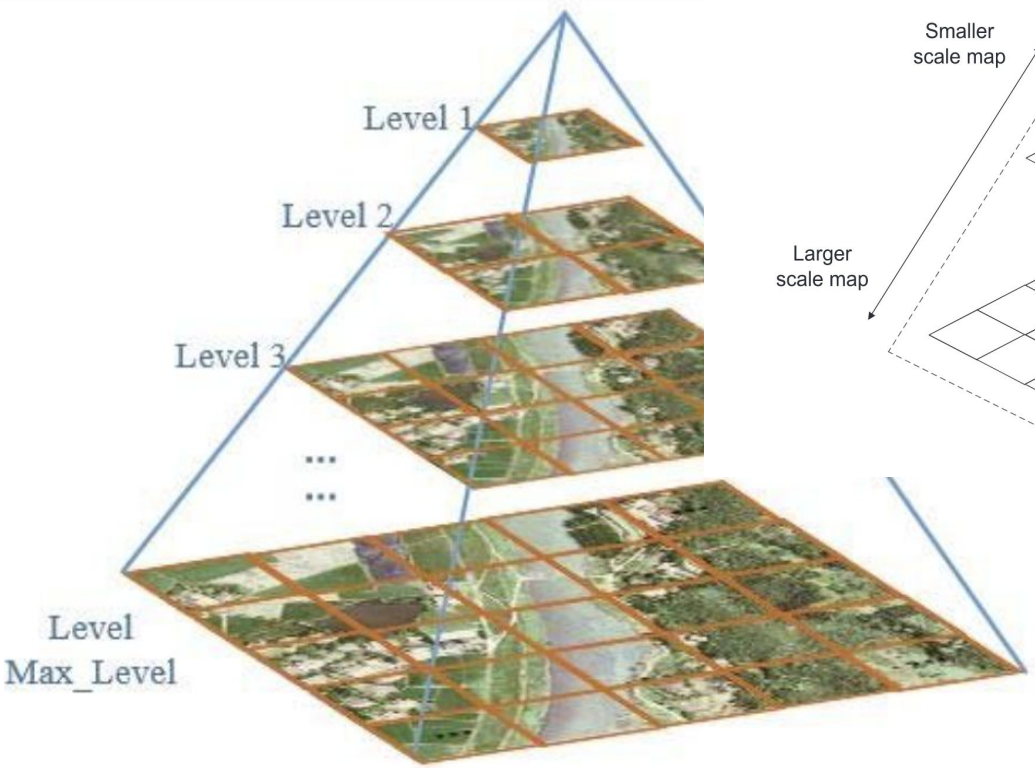
2. Big Data

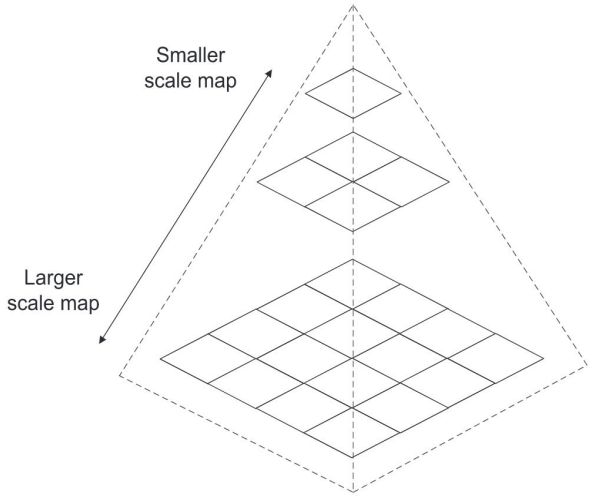


3. Interactivity

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

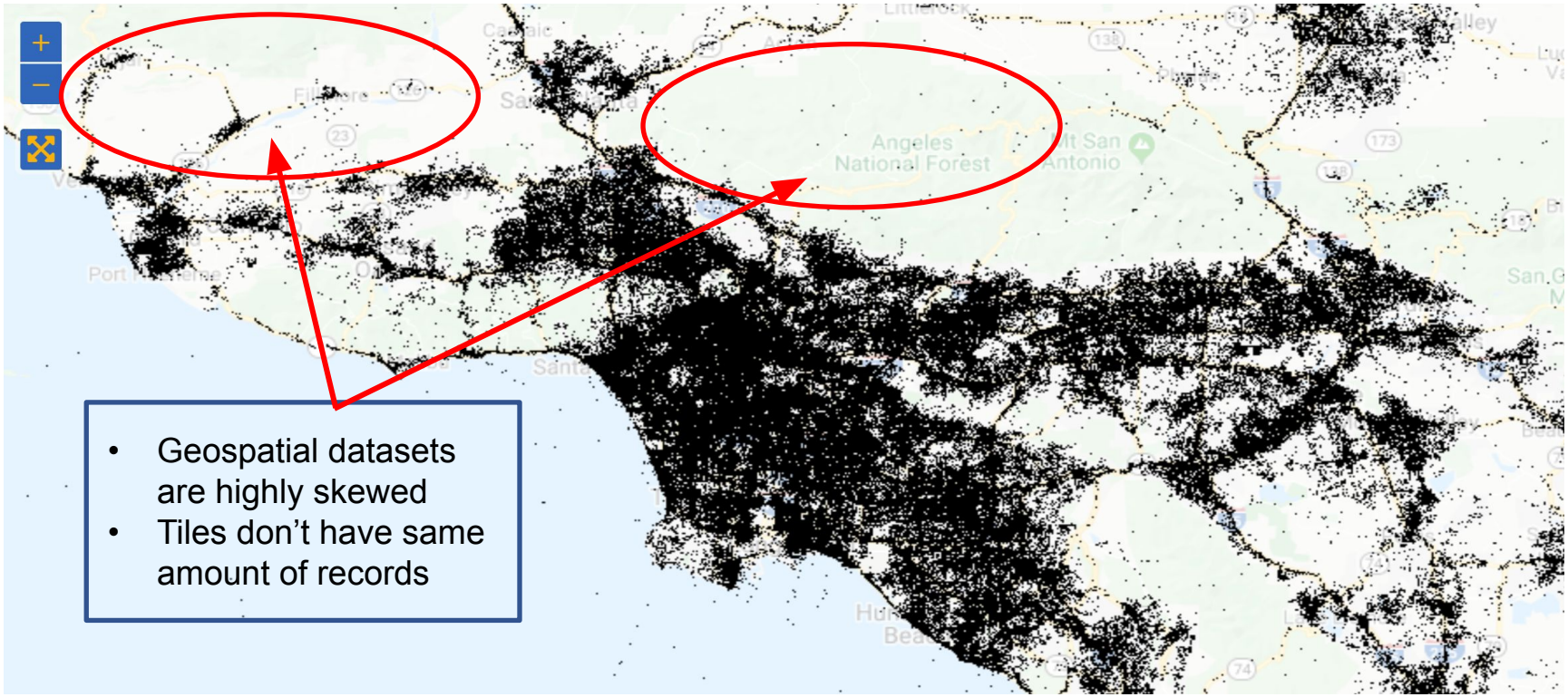
Not scalable for multiple datasets, deeper zoom levels or larger data

Multi-level Visualization using Vector Data

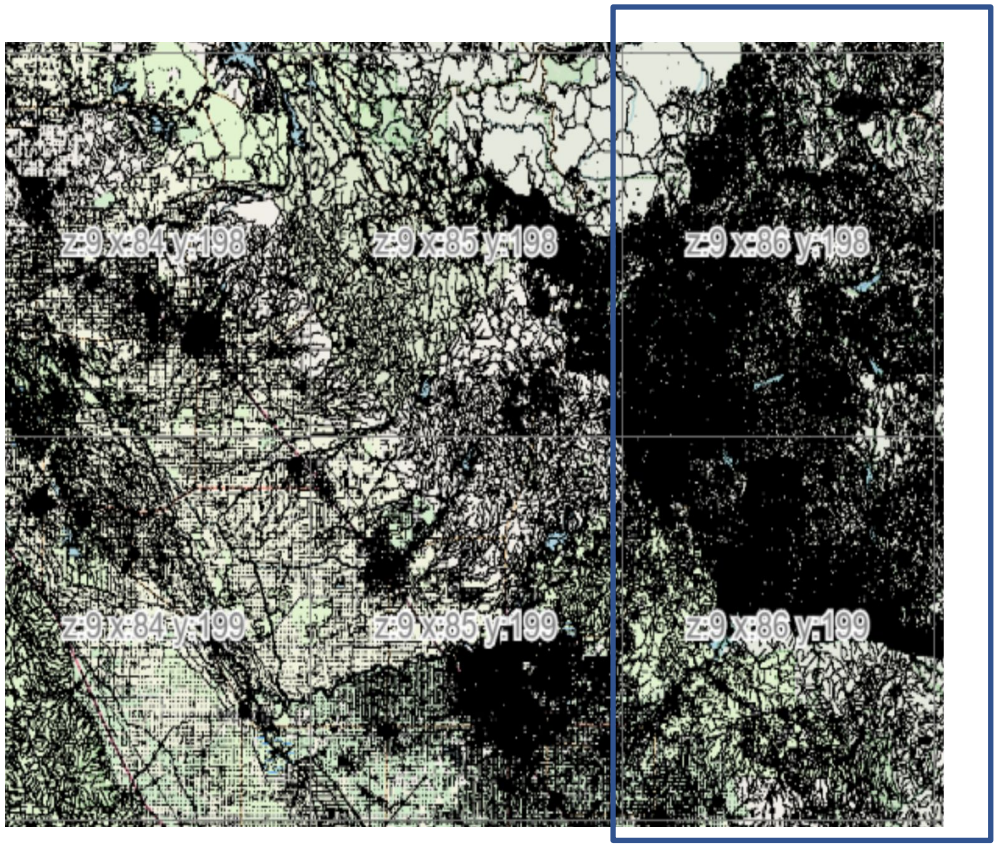
- Vector data: Raw data with attributes like locations (lat, lon), shape, etc
- Images are generated on-the-fly every time a tile is requested by the user

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

Typical Geospatial Datasets

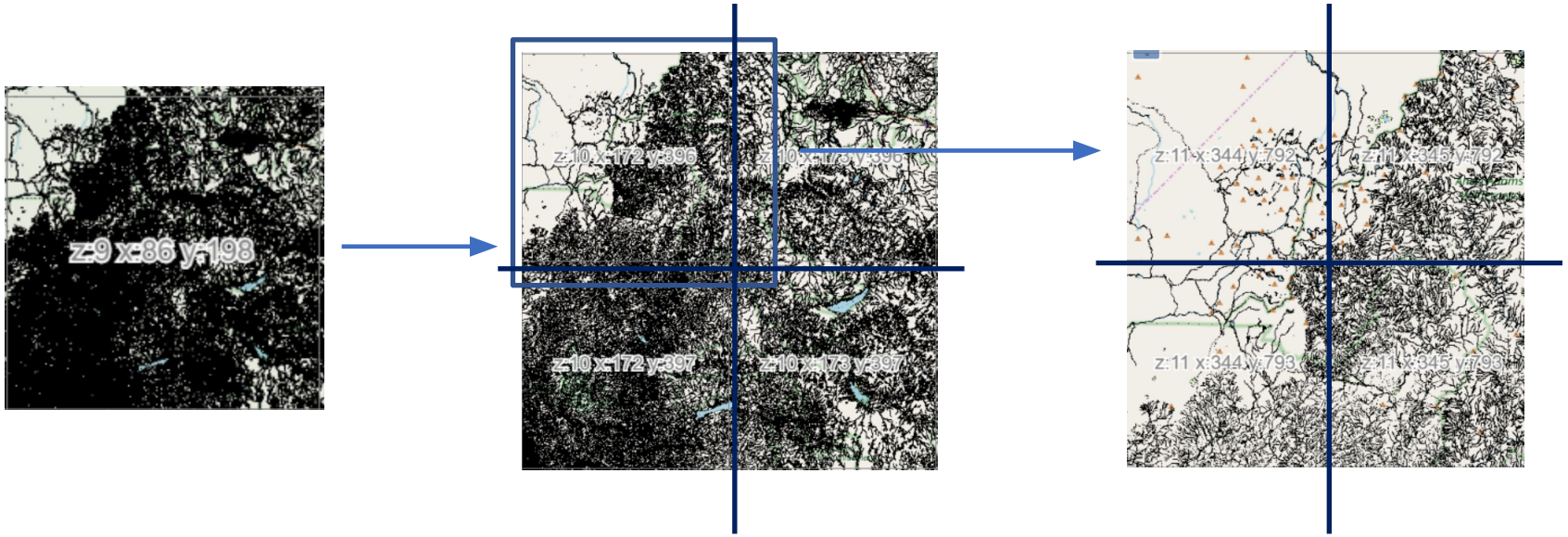


Varying tile density across the area



Tiles has more dense records compared to the remaining four

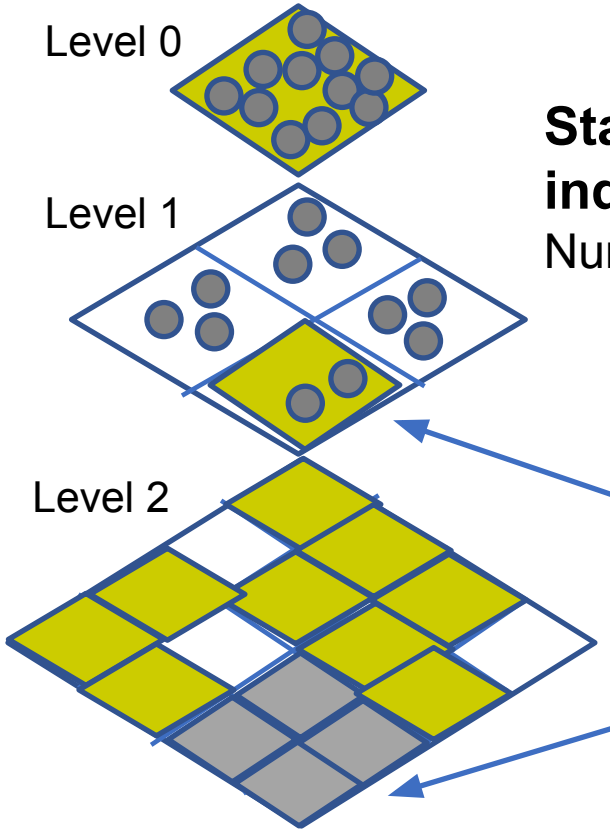
Tile density across zoom levels



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



Standard image indexes
Number of Tiles: 16

AID
Number of Tiles: 14

Let's assume any tile having records <= 2 can be generated on-the-fly

So it becomes a **data tile**

The children tiles become **shallow tiles**

Defining the maximum size of the data tile is given a name **Threshold(θ)**

Threshold for tile classification

A parameter based on the size of each tile

Tile Size > θ → Image Tile

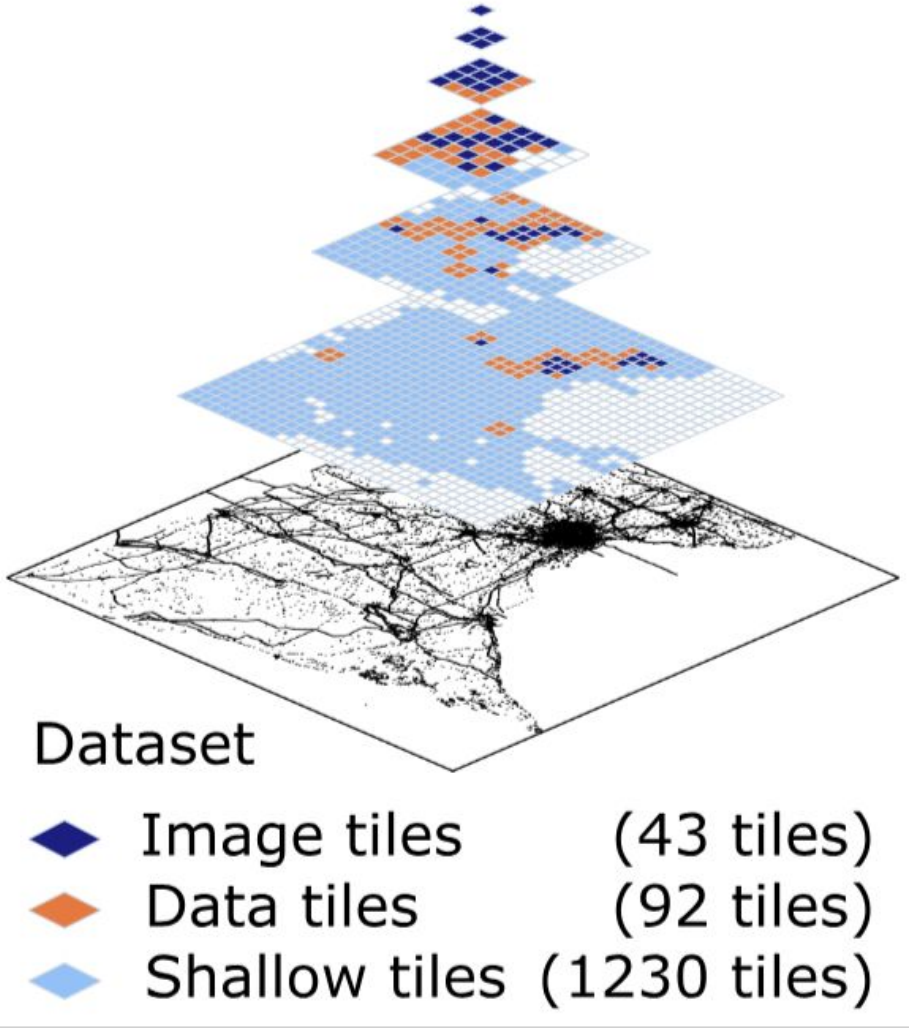
Tile Size <= θ
Parent Tile = Image Tile → Data Tile

Tile Size <= θ
Parent Tile = Data Tile → Shallow Tile

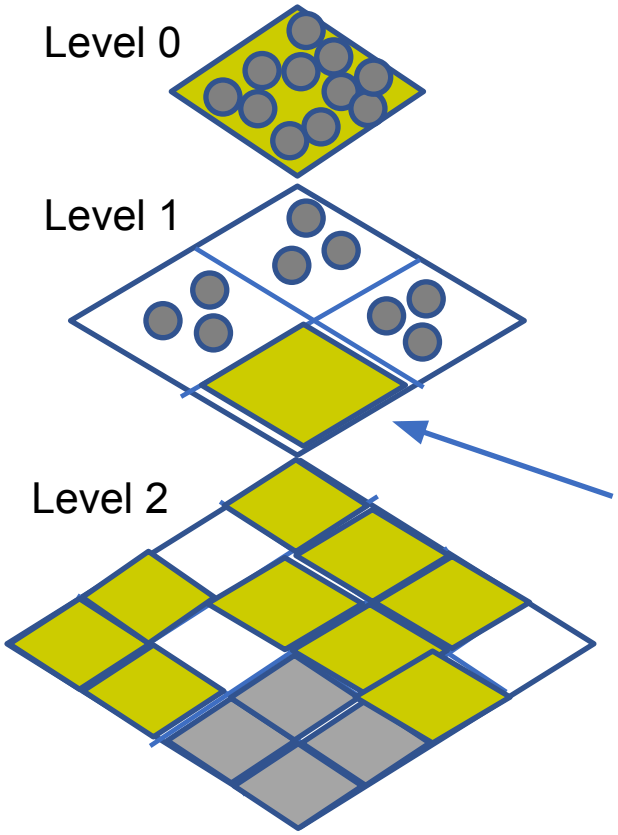
Tile Size = 0 → Empty Tile

Shallow Tile, Empty Tile → No physical storage

AID


Dataset
Image tiles (43 tiles)
Data tiles (92 tiles)
Shallow tiles (1230 tiles)

AID*



AID* does not create or store data tiles

AID
Number of Tiles: 14

AID*
Number of Tiles: 4

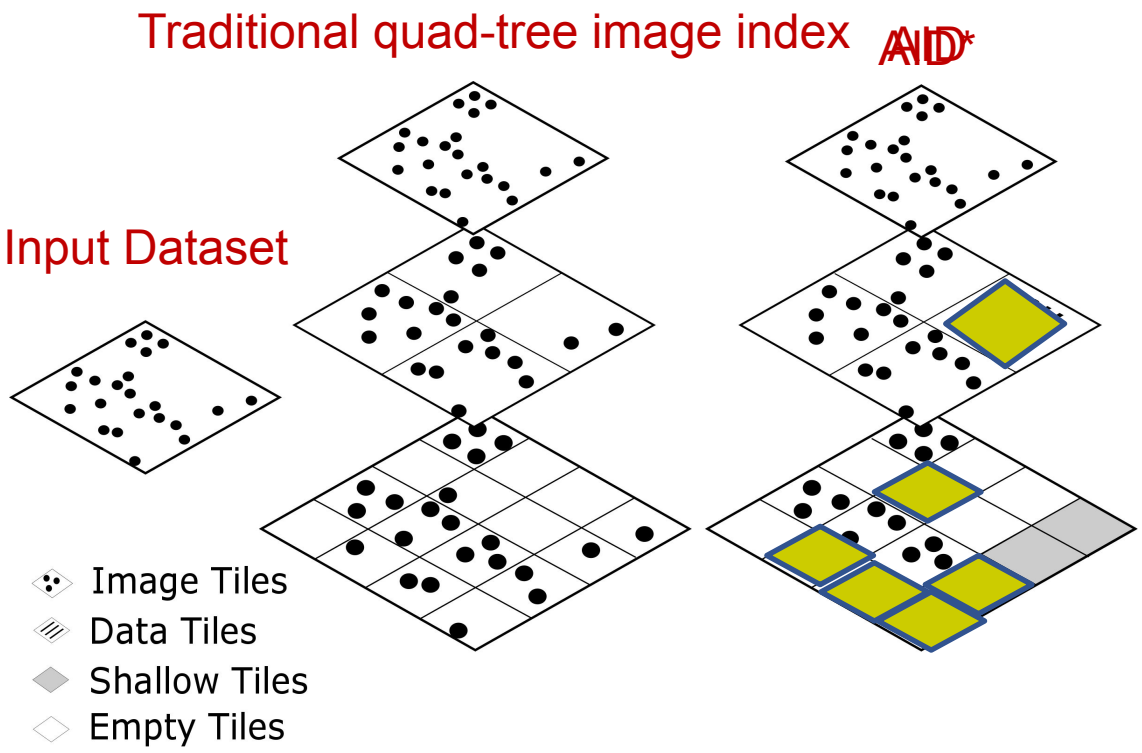
AID*

- 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

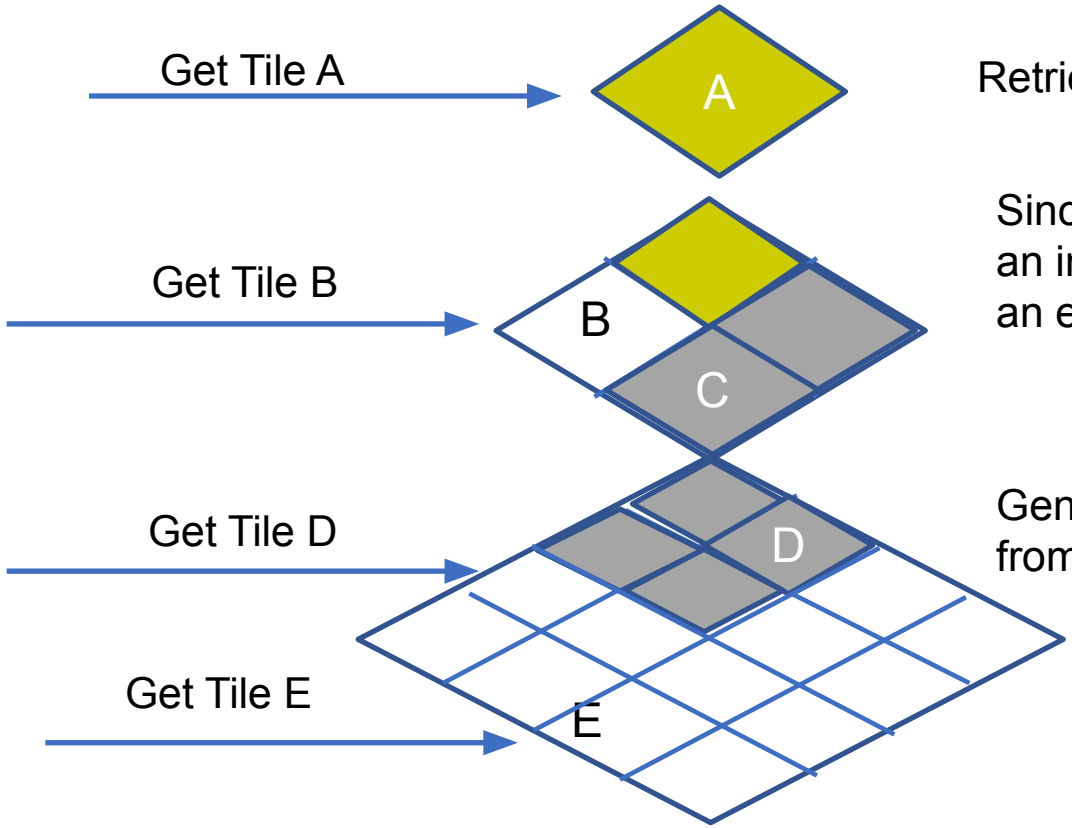
- Data summarization
  - Compute a histogram to summarize the data to calculate the size of each tile
- Tile classification
  - Classify the tiles as image, data, shallow or empty using the histogram
- Tile creation
  - Based on the tile classification image tiles are pregenerated as .png files, data 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*

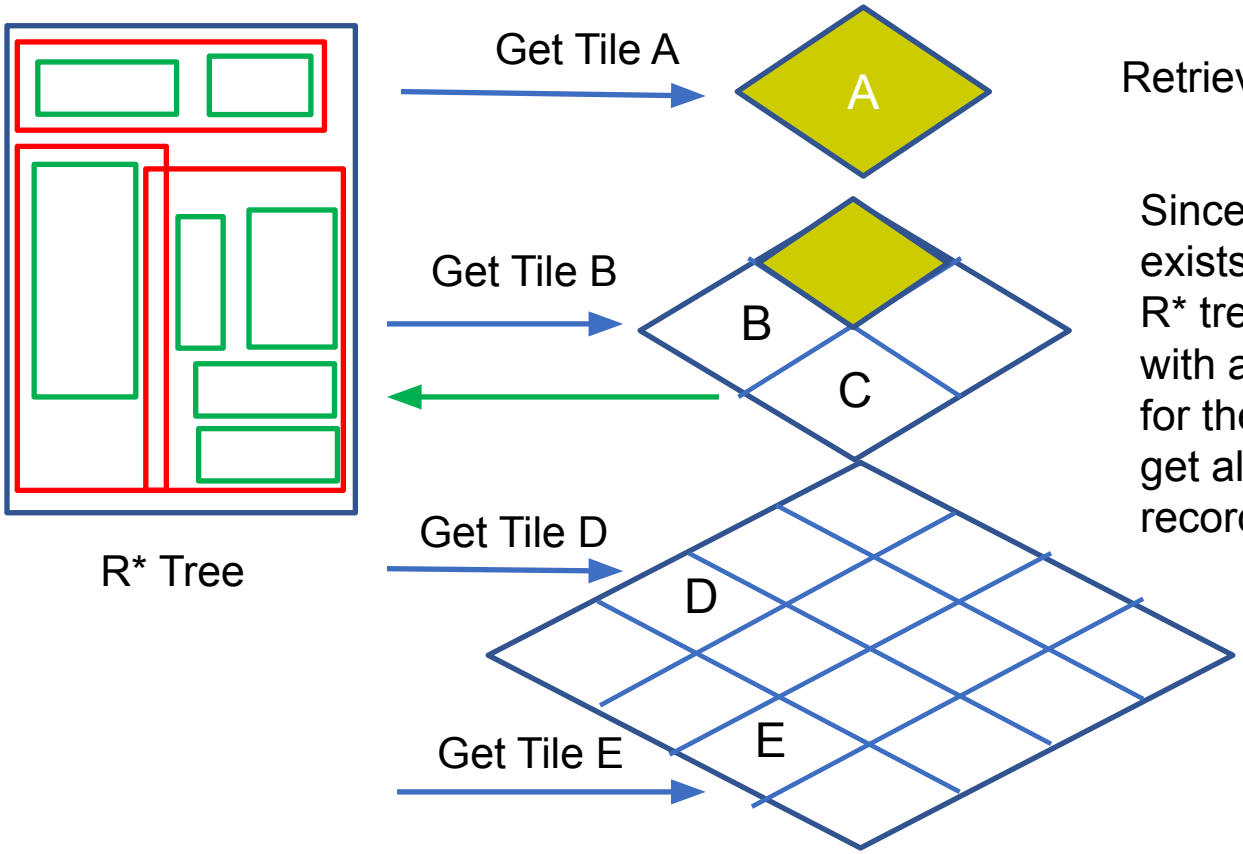


Traditional Index: 15 tiles
AID: 13 tiles
AID*: 8 tiles

Visualization Query for AID (Single-Machine)

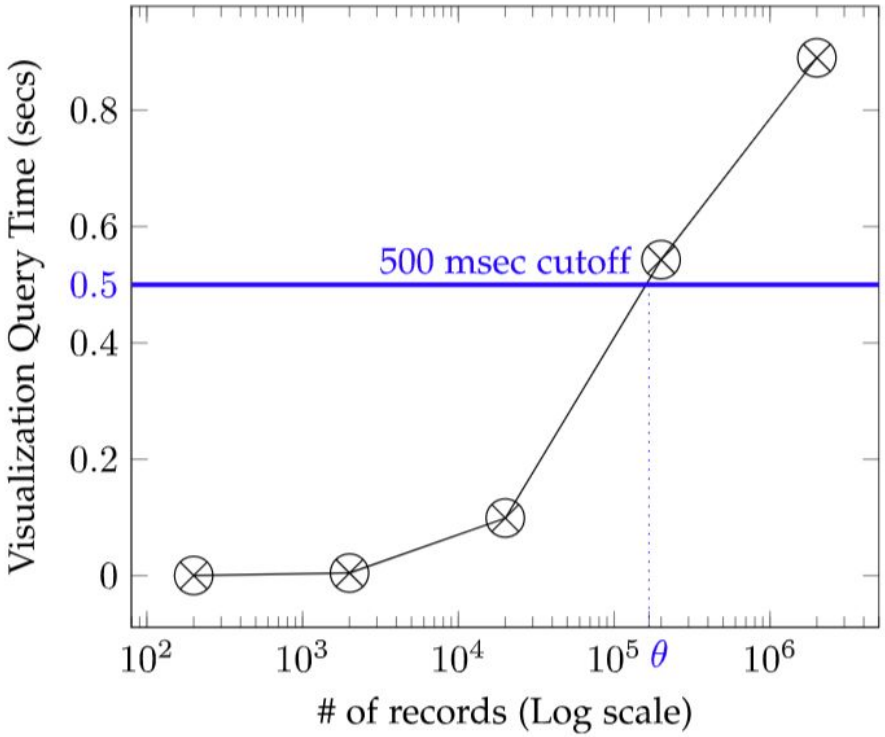


Visualization Query for AID* (Single-Machine)



Tuning Interactivity

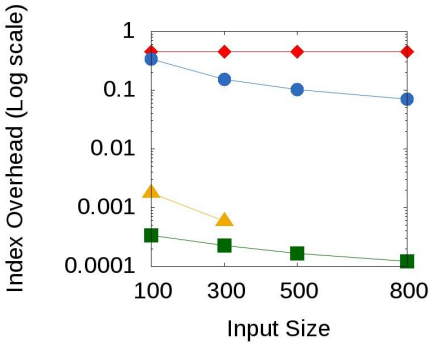
- Through tuning the threshold θ



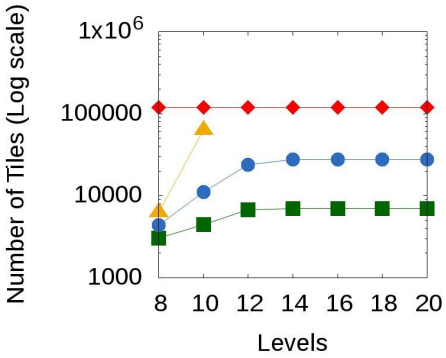
Visualization query time with increasing number of records in a single tile

- Size of θ is the biggest size a data tile can be
- Too big θ => low interactivty
- Too small θ => exponentially growing image index

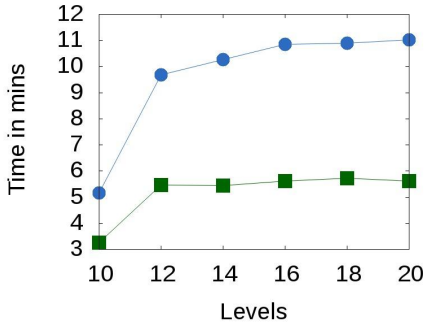
Experiments



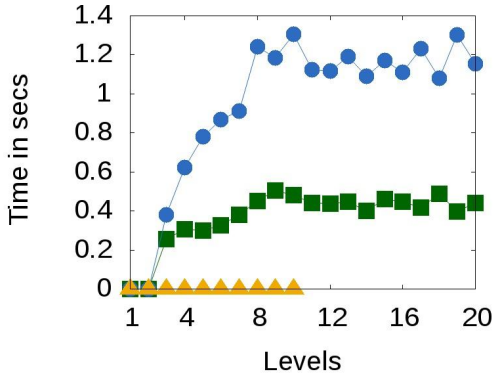
Input Size Vs Index Overhead



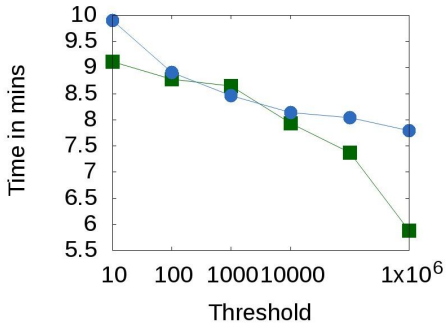
Levels Vs Number of Tiles



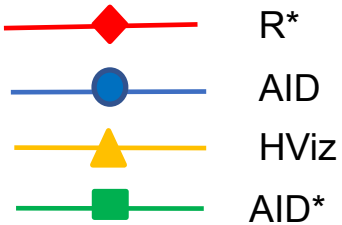
Levels Vs Const Time



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