



ARTS1422 Data Visualization

Lecture 6

Geographical Data Visualization

Quan Li
Spring 2024
2024.03.14

OUTLINE

- 1 Geographic Data
- 2 Map Projections
- 3 Point Data Visualization
- 4 Line Data Visualization
- 5 Region Data Visualization
- 6 Geographic Datasets
- 7 Applications in Urban Problems

Geographic Data

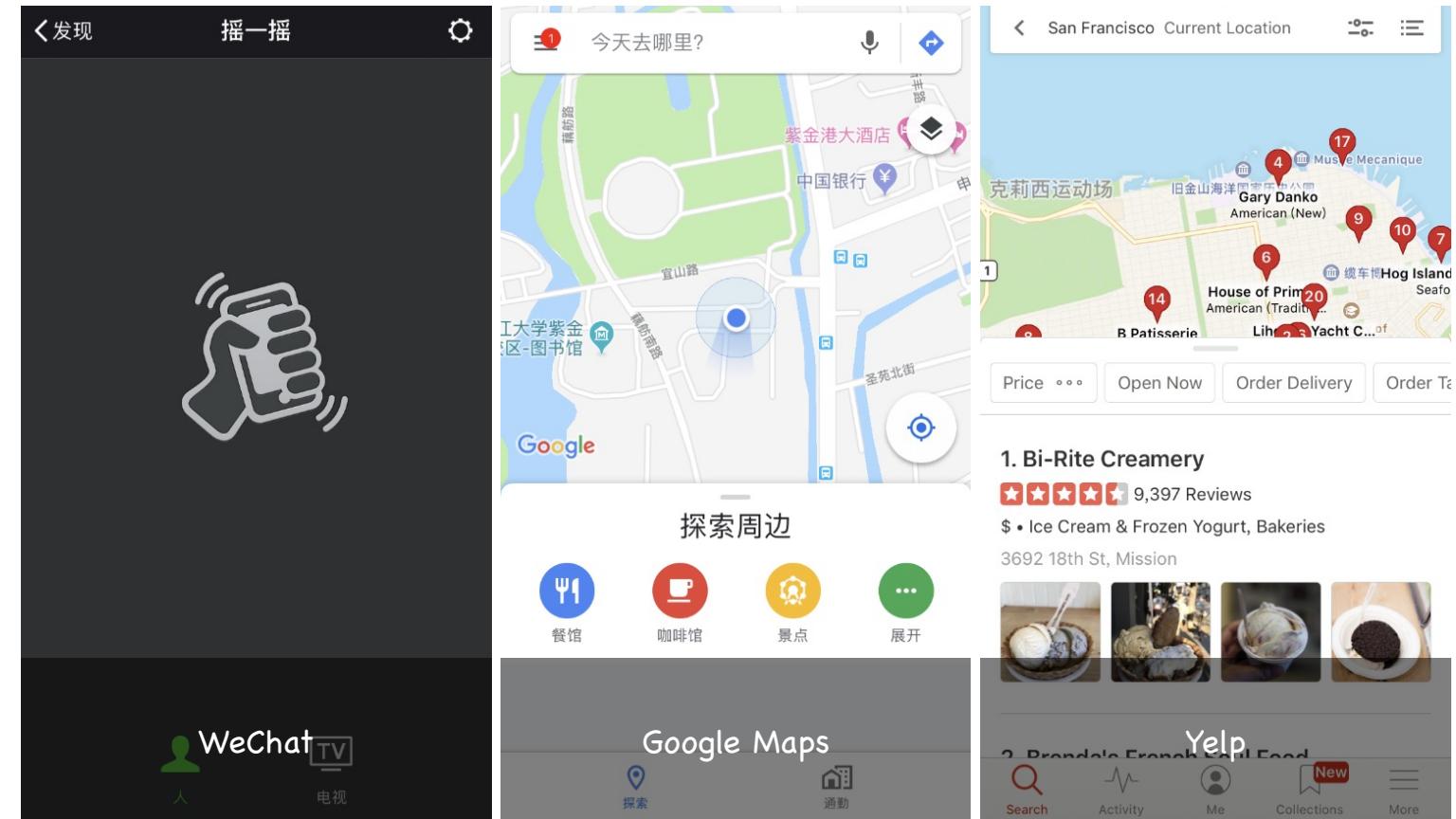
The background features a stylized Earth globe with a blue and teal color palette. The globe is covered in a grid pattern where each square contains a binary digit (0 or 1). A bright, multi-colored sunburst or lens flare effect is positioned in the upper right quadrant, casting light rays across the scene. The overall composition suggests a theme of global digital data.



Geographic Data

- Geographic data, or location-based data describe **the locations of objects**
- Geographic space is the space where human lives, which makes the information carrier and mapping very special and valuable
- Generated by mobile devices and sensors everyday

Location- Based Services



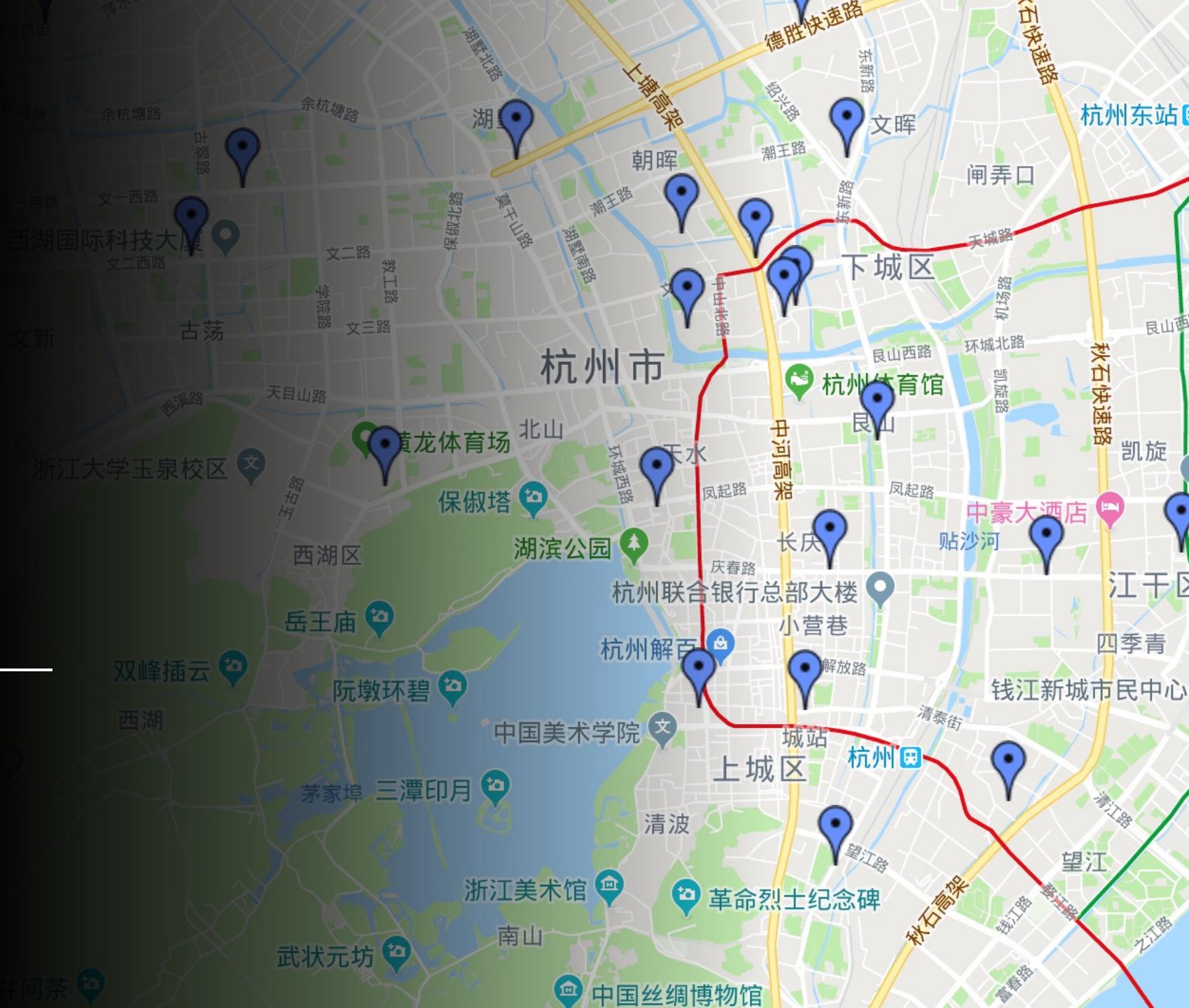
Point Data Visualization



Point Data

- Points scattered in the geographical space, with longitude and latitude information, optionally with weights.
 - Landmarks on the map
 - Restaurants of your neighborhood
- Marking the data on the map would be the most direct and simplest way to visualize point data.
- Point-based visualizations are widely used.

Supermarkets in Hangzhou



Temporal Point Data



上海科技大学
ShanghaiTech University

1690

1 publication



STANFORD UNIVERSITY
THE BILL LANE CENTER
FOR THE AMERICAN WEST

Rural West Initiative, Mapping Journalism's Voyage West.

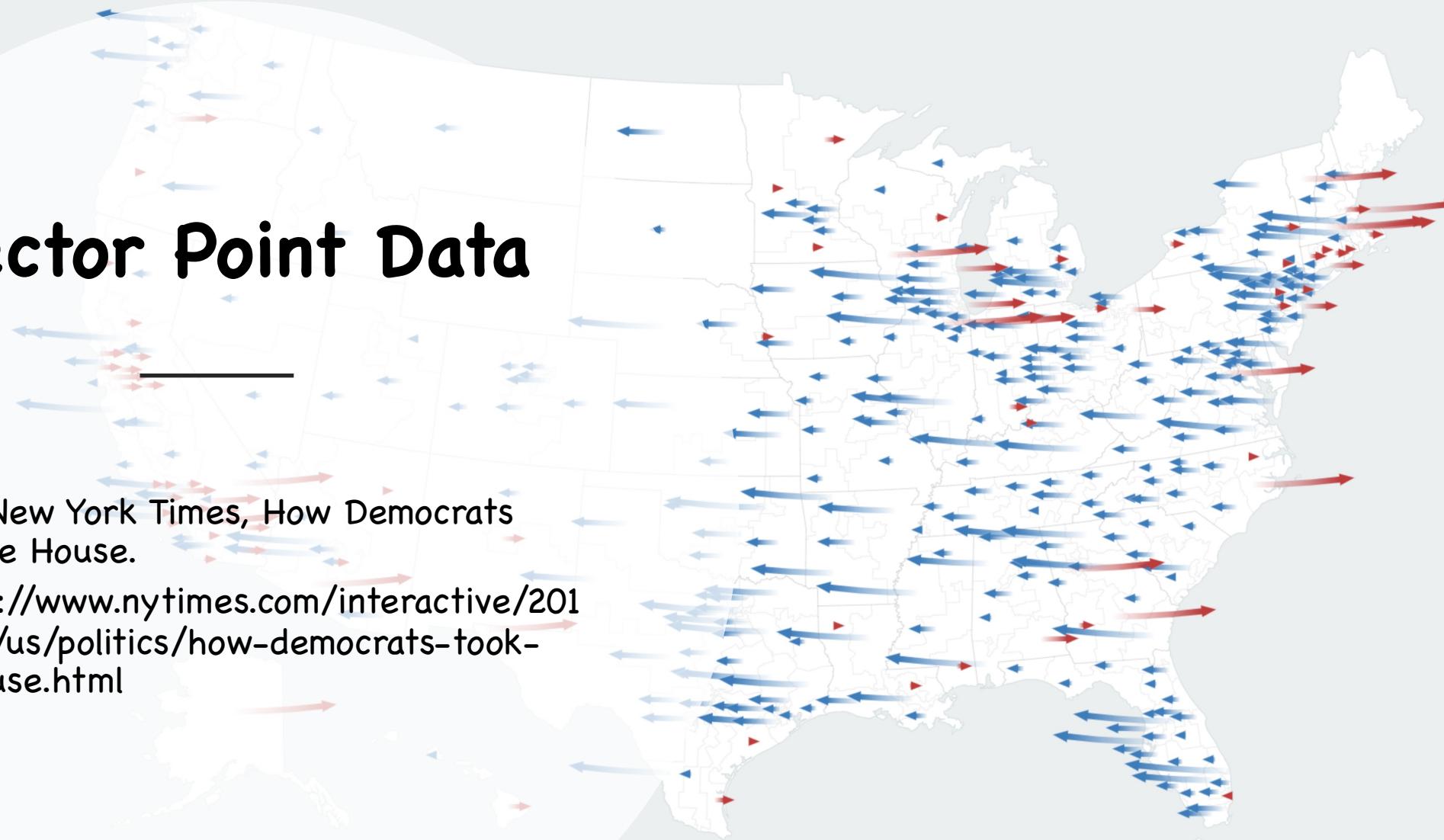
https://web.stanford.edu/group/ruralwest/cgi-bin/drupal/visualizations/us_newspapers



立志成才报国裕民

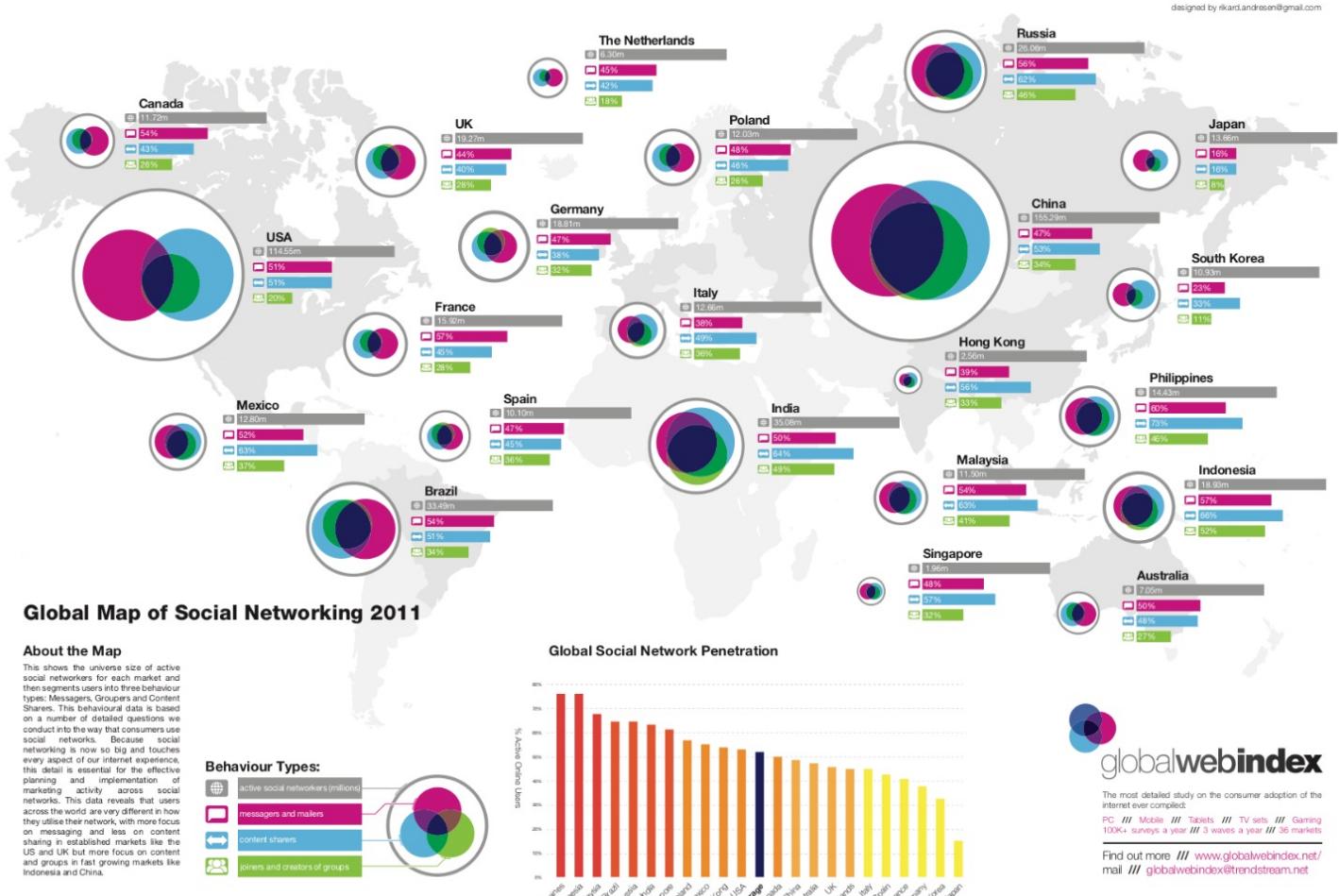
Vector Point Data

- The New York Times, How Democrats Took the House.
- <https://www.nytimes.com/interactive/2018/11/07/us/politics/how-democrats-took-the-house.html>



Global Map of Social Networking in 2011

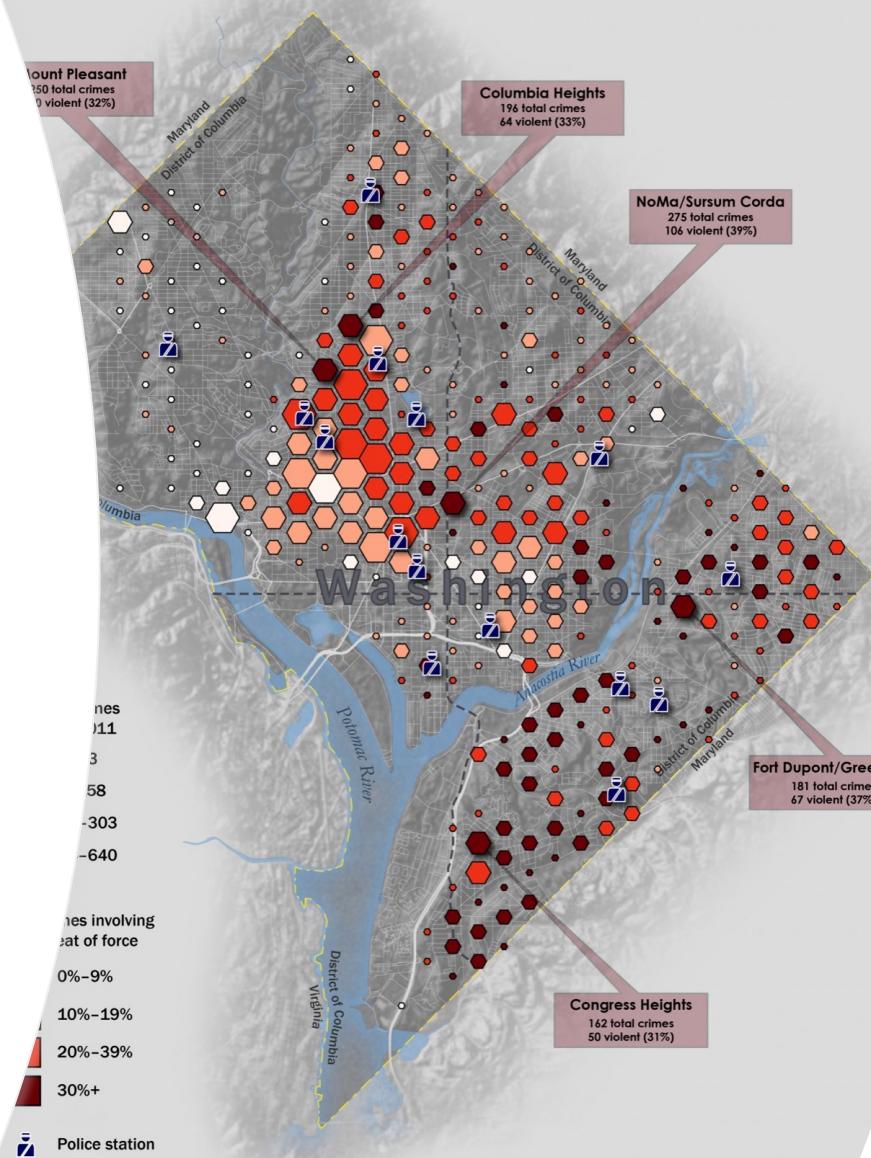
- Global Web Index, Global Map of Social Networking.
- <https://blog.globalwebindex.com/chart-of-the-day/new-globalwebindex-infographic/>



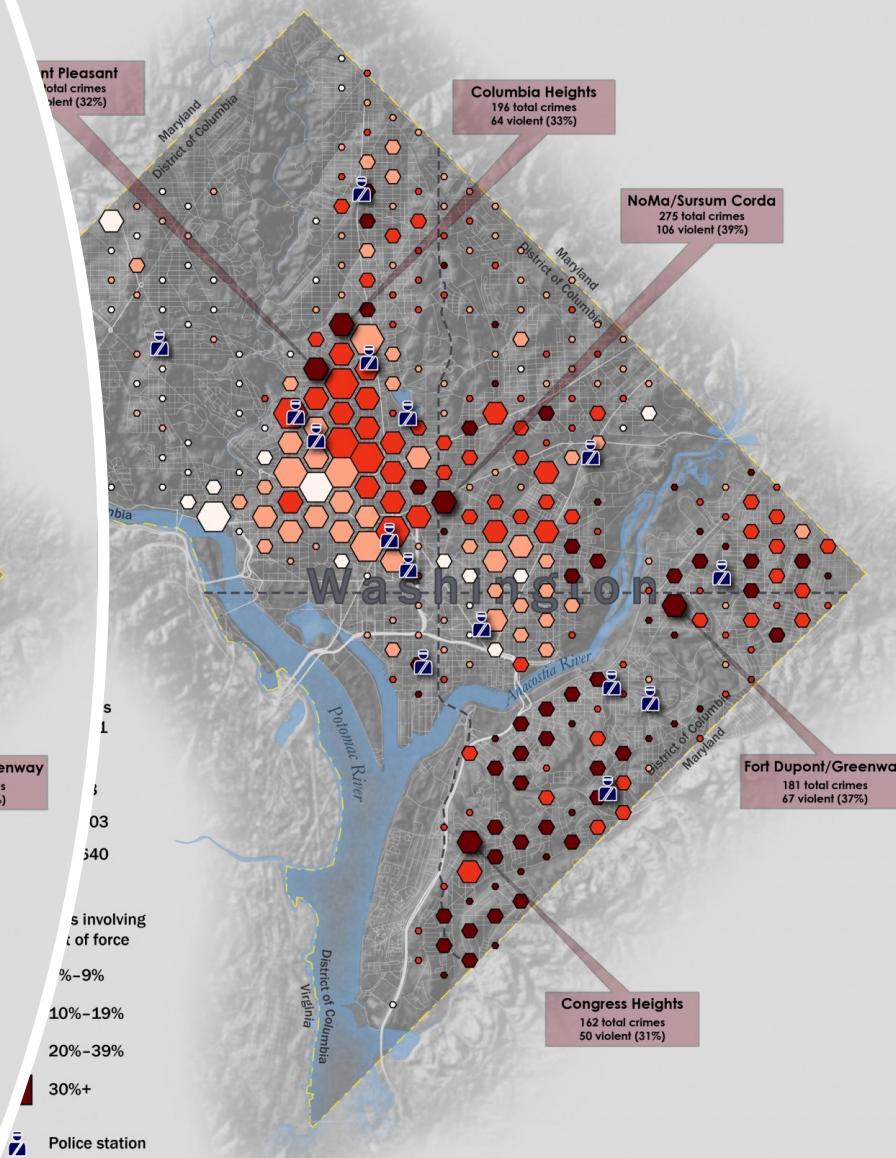
Mapping Violent Crime in Washington, D.C.

Violent Crime in Washington, D.C.

- Cooper Thomas, Violent Crime in Washington, D.C.
- <https://visual.ly/community/infographic/geography/violent-crime-washington-dc>



Source: Metropolitan Police Department 2011 Crime Incident Statistics



Source: Metropolitan Police Department 2011 Crime Incident Statistics

Every death on every road in Great Britain

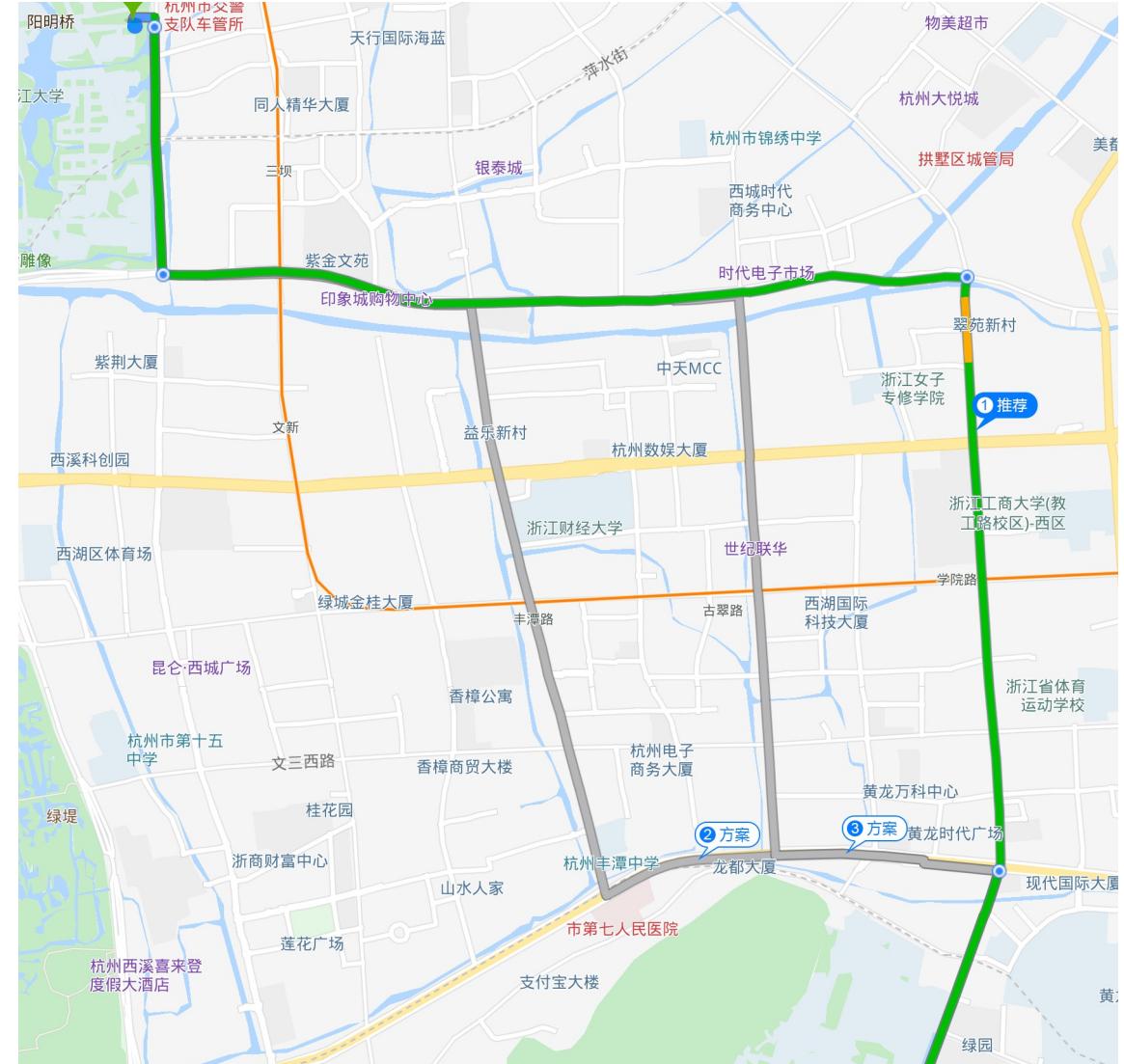
- BBC, Every death on every road in Great Britain 1999–2010.
- <https://www.bbc.co.uk/news/uk-15975724>



Line Data Visualization

Line Data

- Line data comprise locations and the paths between them.
- Each line has a length property, namely the distance between two locations.



Air Traffic Network

- Martin Grandjean, Connected World: Untangling the Air Traffic Network.
- <http://www.martingrandjean.ch/connected-world-air-traffic-network/>



Facebook Friend Relationship

- Facebook, 2013.
- <https://bit.ly/2P7WZep>

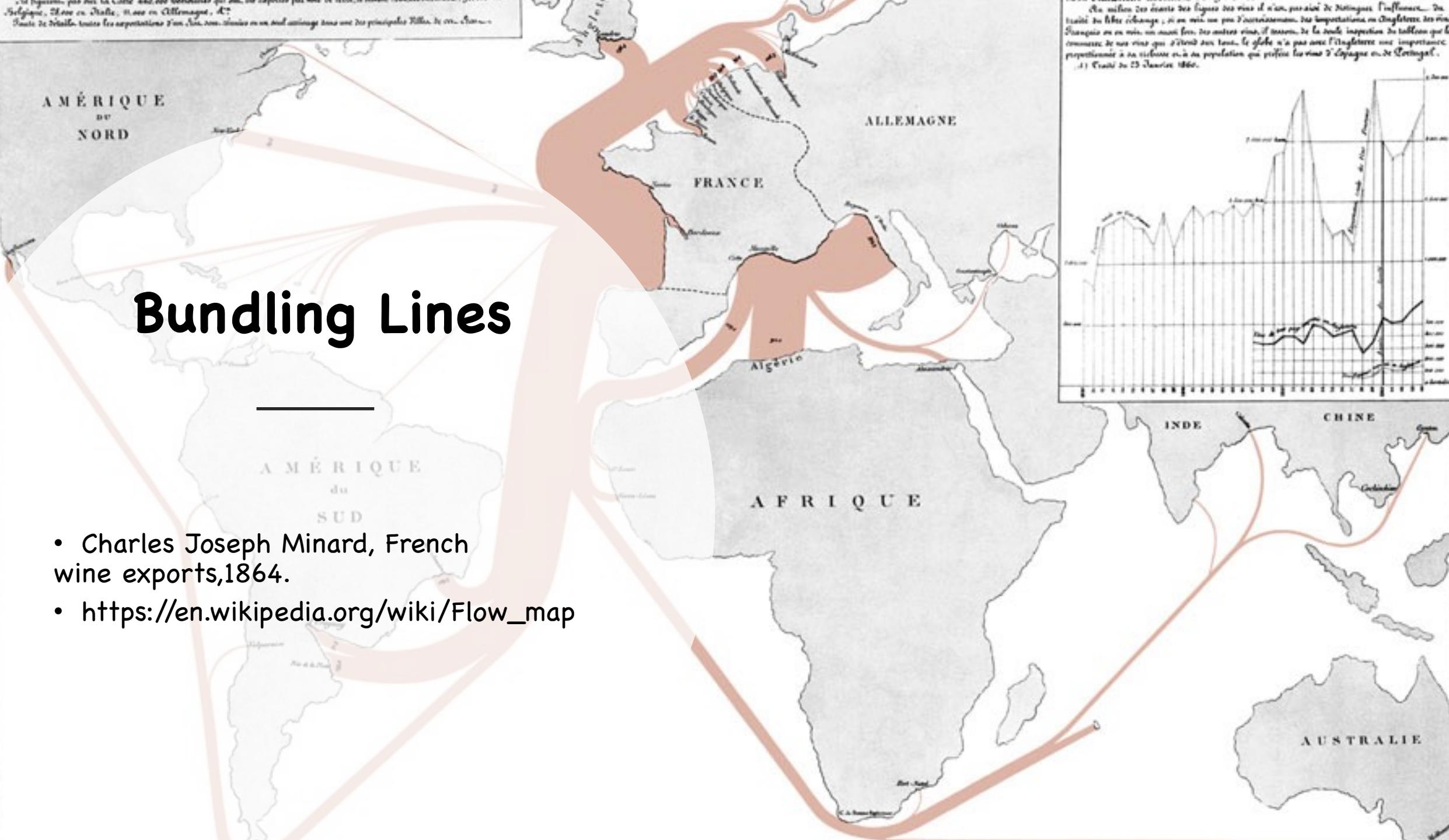
...Belgique, Italie en Italie, Italie en Allemagne, etc.
Toute la détaillée toutes les exportations d'un pays sont résumées en un seul chiffre dans une des principales villes de ce pays.

On sait des étoiles des lignes des vins et c'est parmi de distingués l'influence... du traité de libre échange, si on veux, sur l'exportation des importations en Angleterre des vins français ou en tout cas pour ces autres vins, il existe de la toute importance de tellement que le commerce de nos vins qui s'étend sur toute la globe n'a pas avec l'Angleterre une importance proportionnelle à sa richesse ou à sa population qui préfère les vins d'Espagne ou de Portugal.

Le traité de 25 Janvier 1860.

Bundling Lines

- Charles Joseph Minard, French wine exports, 1864.
- https://en.wikipedia.org/wiki/Flow_map

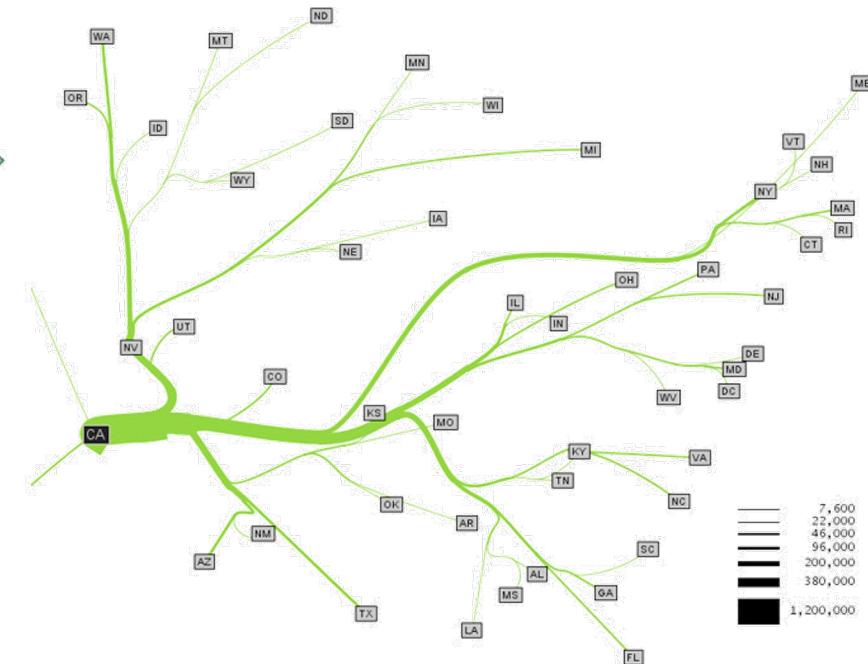


Flow Map

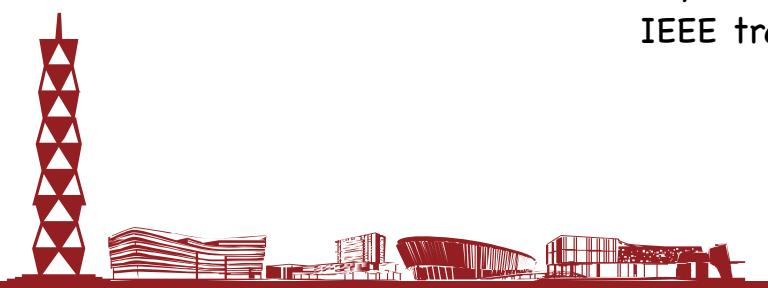


Linking the immigrations from California
to other states with arrows

Buchin, Kevin, Bettina Speckmann, and Kevin Verbeek. "Flow map layout via spiral trees." IEEE transactions on visualization and computer graphics 17.12 (2011): 2536-2544.

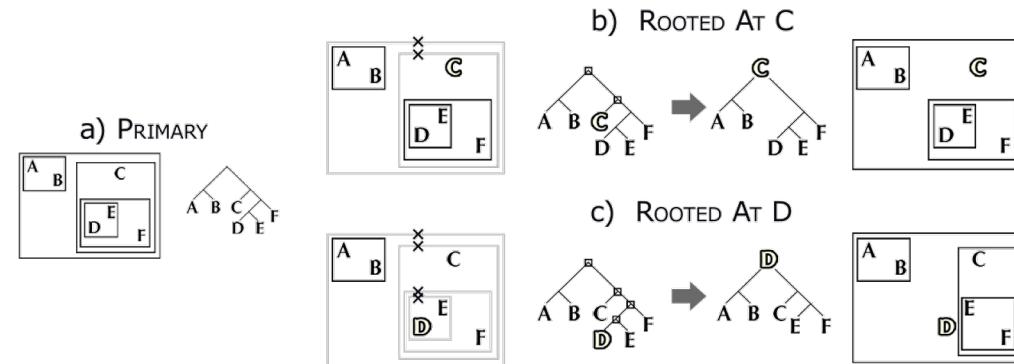


Visualized with Flow Map

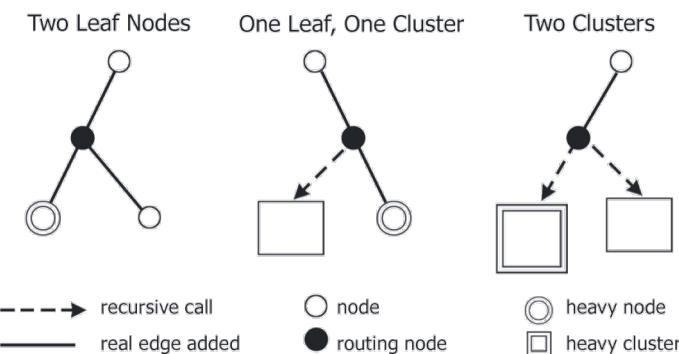


Flow Map Procedures

- Computing the bundles.



- Layout of the lines.



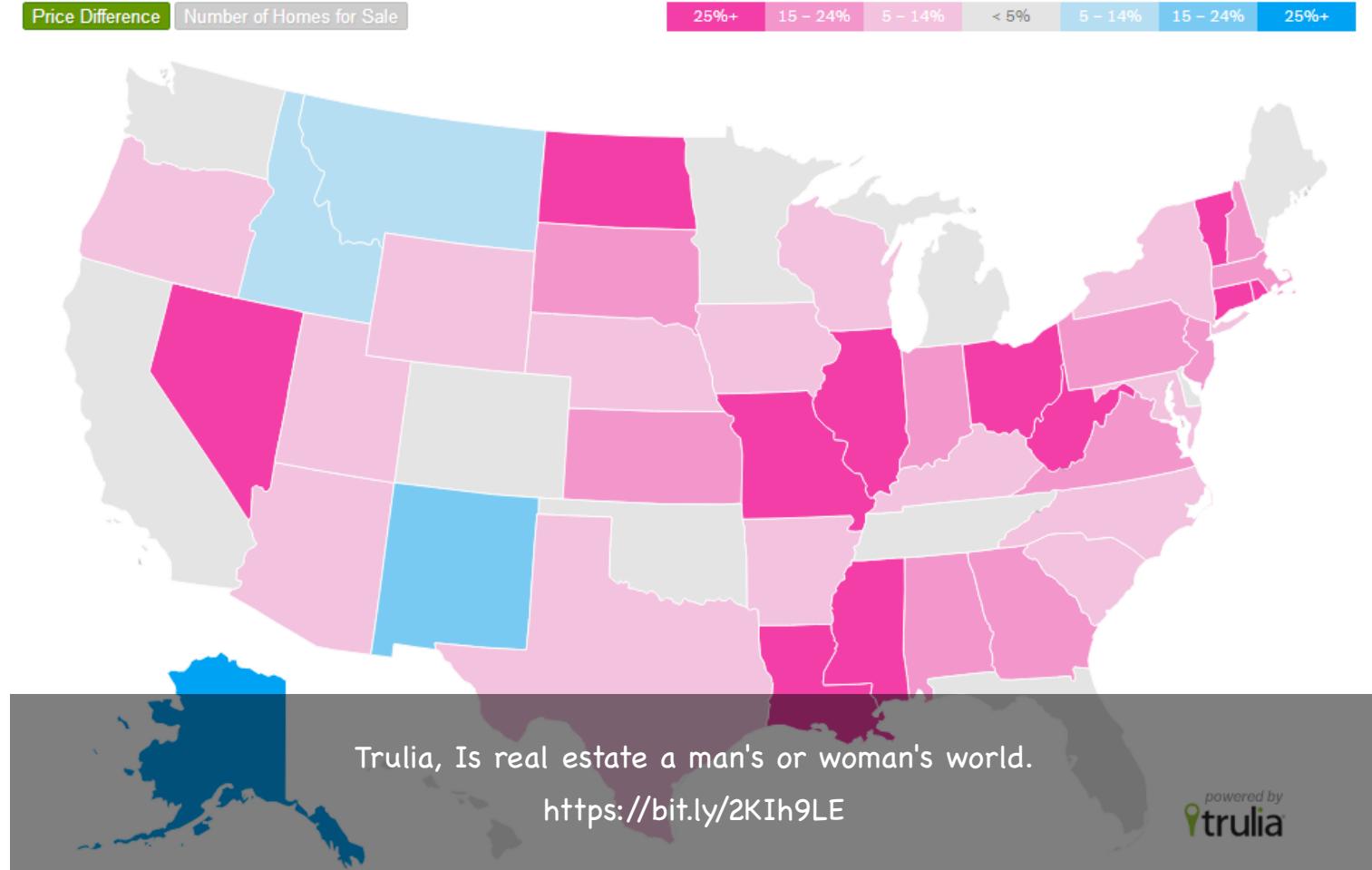
Buchin, Kevin, Bettina Speckmann, and Kevin Verbeek. "Flow map layout via spiral trees." IEEE transactions on visualization and computer graphics 17.12 (2011): 2536-2544.

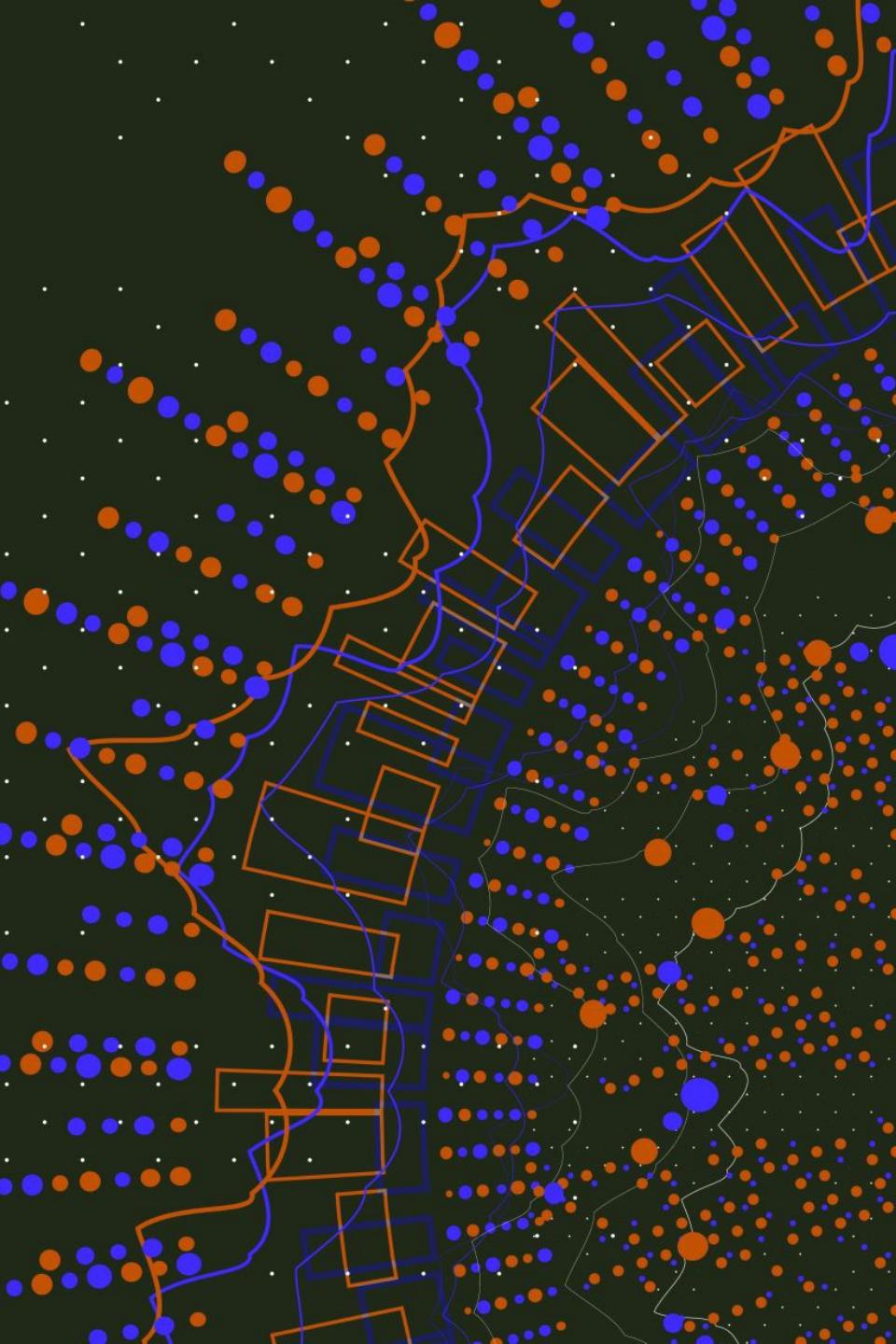


A close-up photograph of a television screen, showing the underlying RGB pixel structure. The screen is composed of a grid of small, colored dots: red, green, and blue. These dots are arranged in a repeating pattern across the entire frame, creating a textured, grid-like appearance.

Region Data Visualization

Selling Real Estate: Men vs. Women



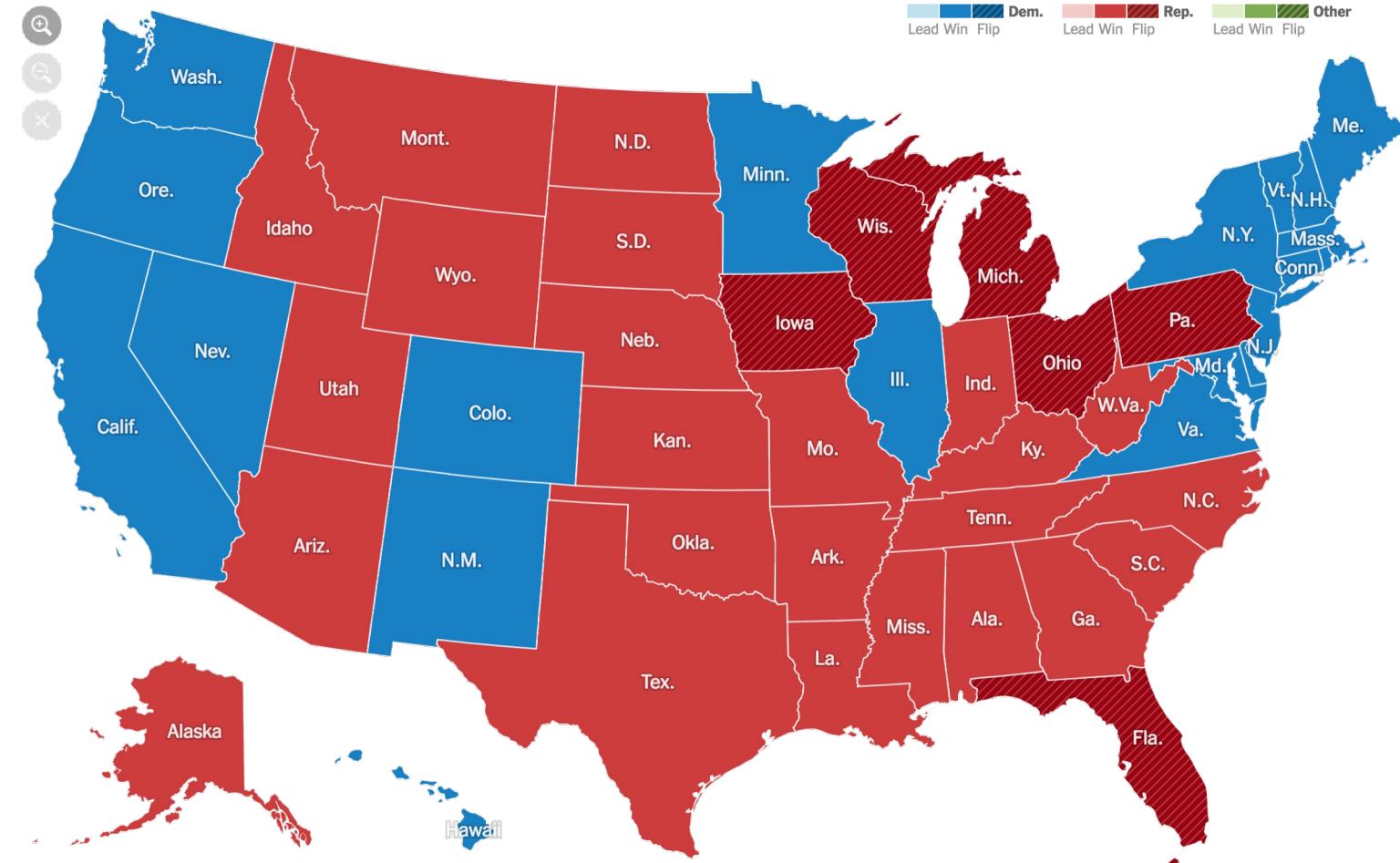


Choropleth Map

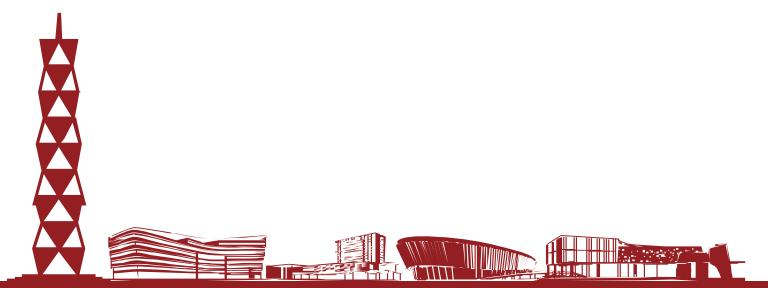
- A thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map.
- Assumes that the data is evenly distributed in the area.
- Represents the scales of data with color.



Choropleth Map



The New York Times, 2016 Presidential Election Results.
<https://www.nytimes.com/elections/2016/results/president>

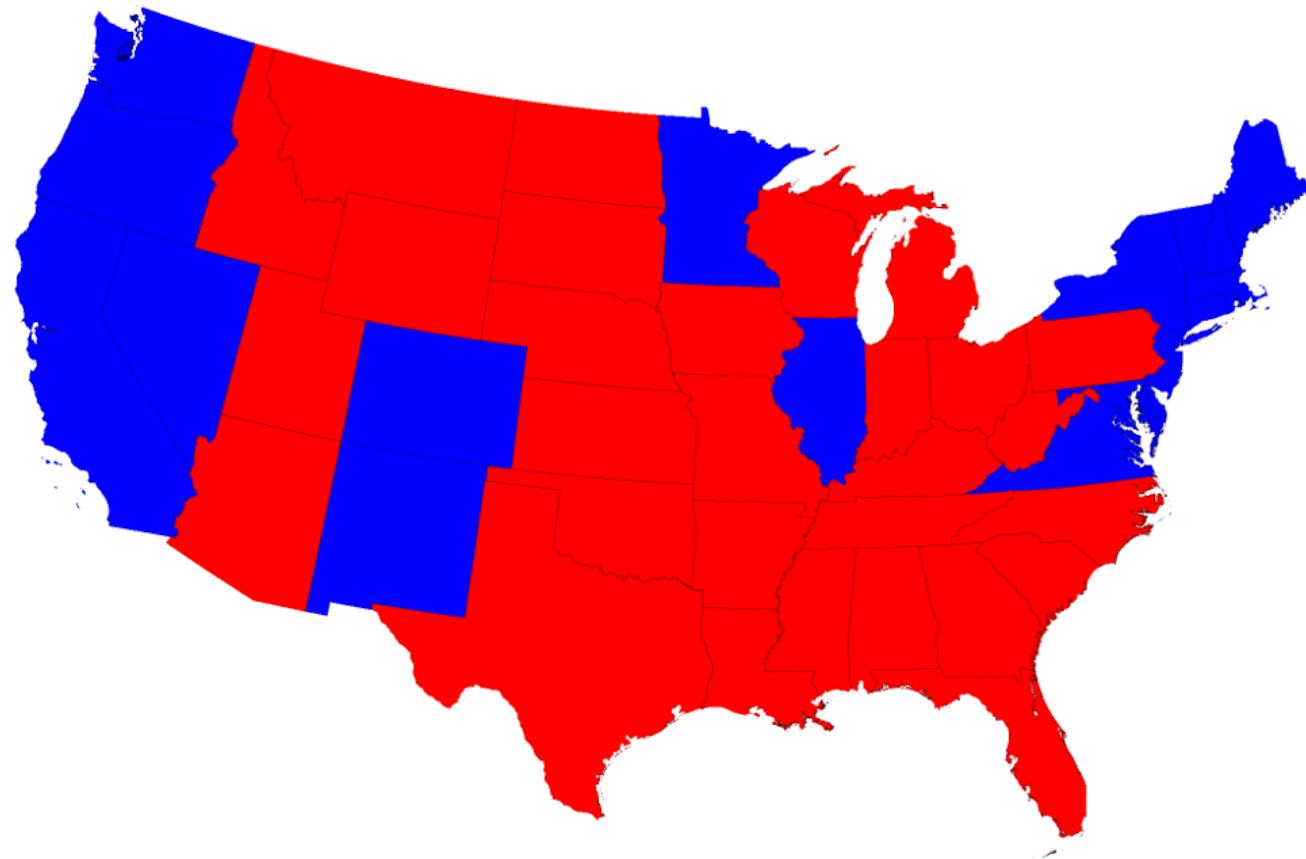


Misleading of Choropleth Map

- The map fails to take account of **the population distribution**.
 - It fails to allow for the fact that the population of the red states is on average significantly lower than that of the blue ones. The blue may be small in area, but they represent a large number of voters, which is what matters in an election.
- When data is accumulated at the place where display space is less, there comes a mismatch, and vice versa.
- This mismatch would very possibly **mislead** the audience.

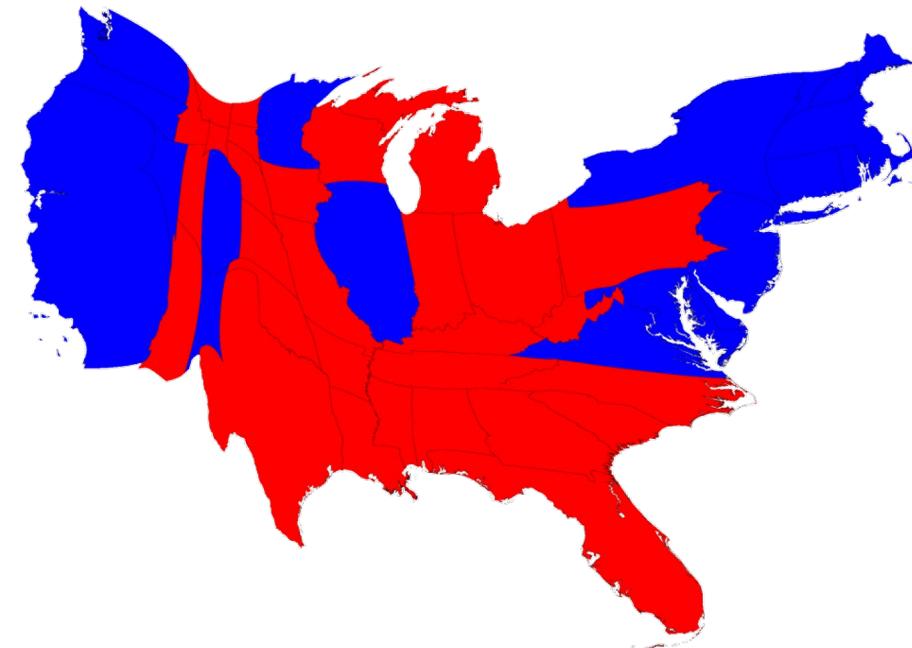
Misleading of Choropleth Map

- Mark Newman, 2016 Maps of the 2016 US presidential election results.
- <http://www-personal.umich.edu/~mejn/election/2016/>



Cartogram Map

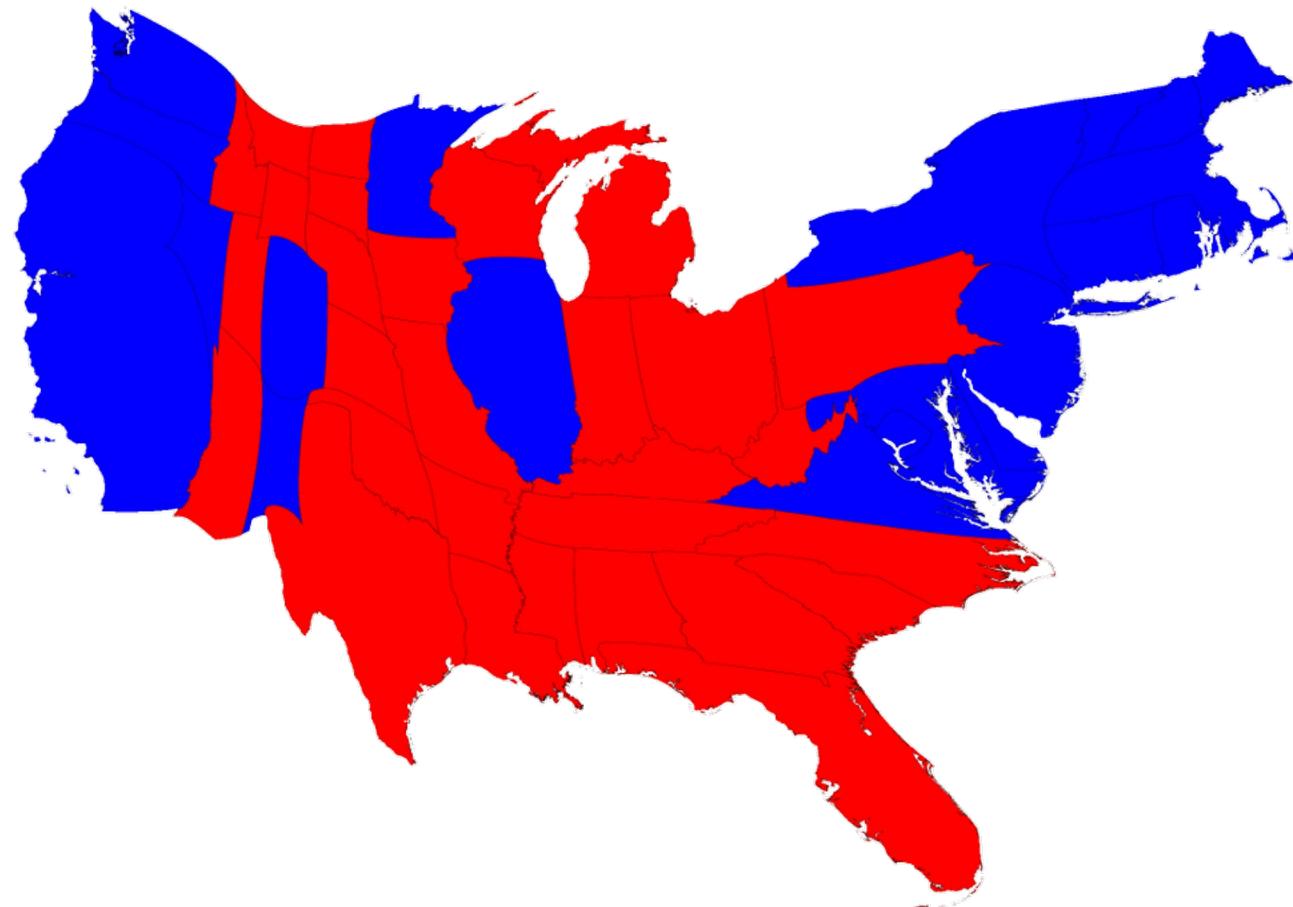
- A map in which some thematic mapping variable – such as travel time, population, or Gross National Product – is substituted for land area or distance.



Mark Newman, 2016 Maps of the 2016 US presidential election results.
<http://www-personal.umich.edu/~mejn/election/2016/>



Cartogram Map



Mark Newman, 2016 Maps of the 2016 US presidential election results.
<http://www-personal.umich.edu/~mejn/election/2016/>



Other Maps



Regular Shape Map

- Visualizing areas with simple shapes like rectangles or circles, so that users can compare the size of data easier.



- The outcome of the UK Brexit EU-referendum is displayed.
- The region areas represent the number of the electorates.
- The colors are indicating the outcome of the referendum (blue: remain/red: leave; the lower the color intensity the closer is the outcome to 50%:50%).

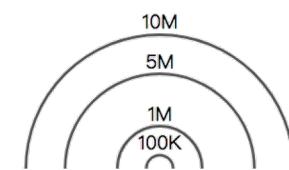
Christian Panse, Rectangular Statistical Cartograms in R: The `recmap` Package.
Journal of Statistical Software



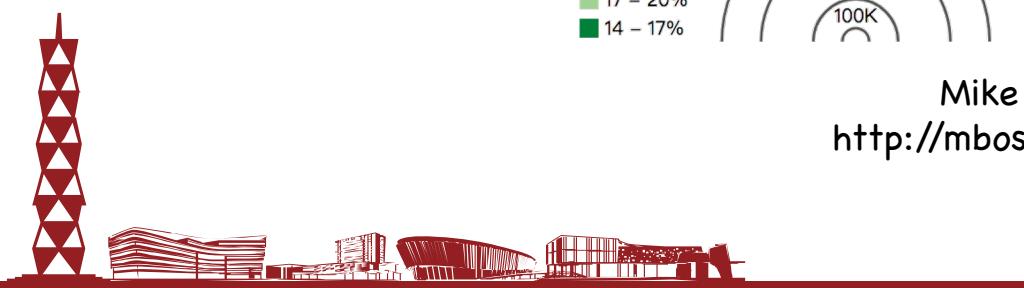
Dorling Cartogram



■ 32 – 35%
■ 29 – 32%
■ 26 – 29%
■ 23 – 26%
■ 20 – 23%
■ 17 – 20%
■ 14 – 17%

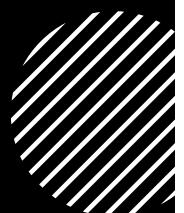


Mike Bostock , Obese People Per State.
<http://mbostock.github.io/protovis/ex/cartogram.html>



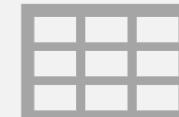
立志成才报国裕民

Multi-Set Map



To represent different attributes of the locations.

restaurants,
gas stations...

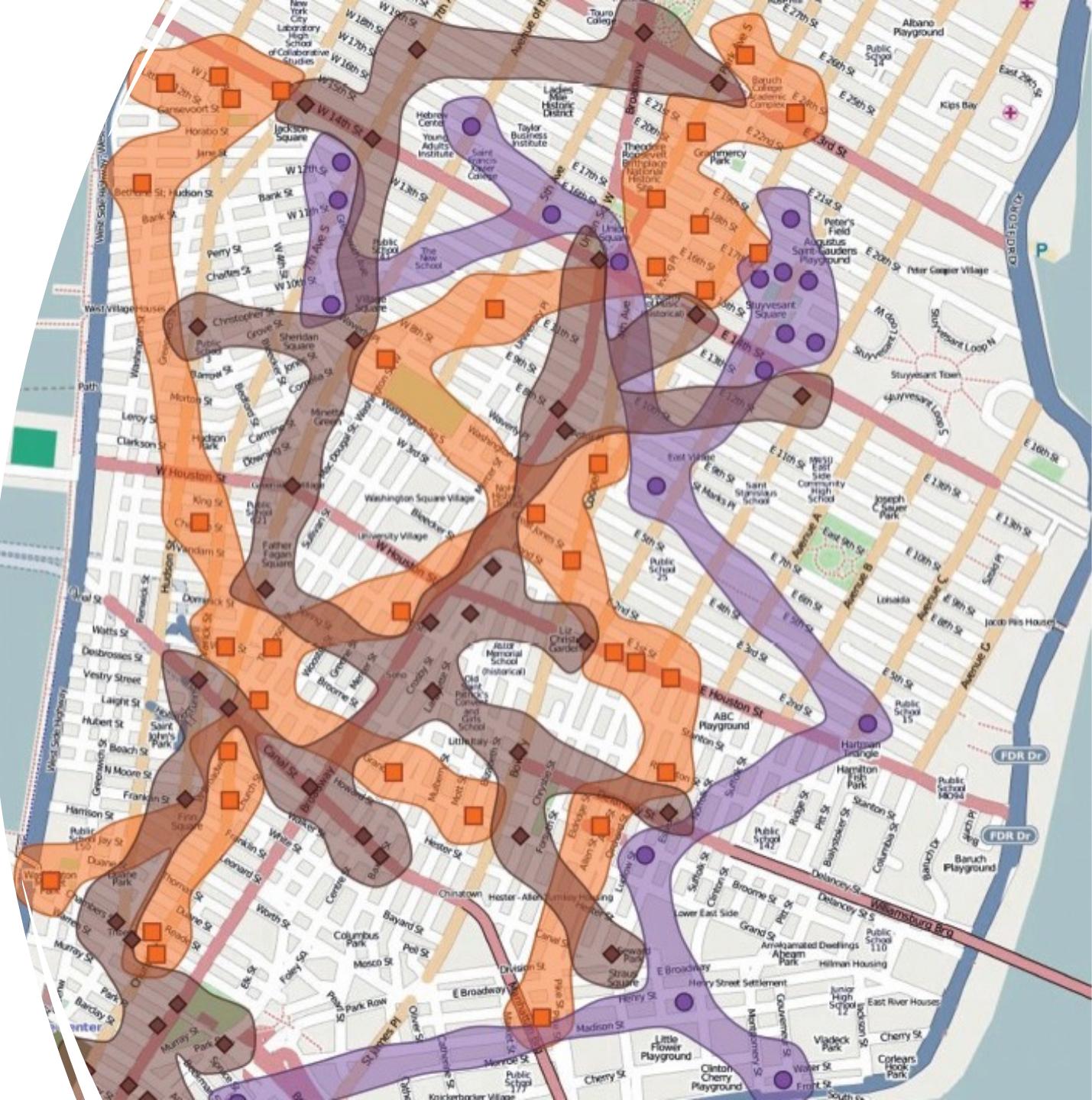


Multiple set representations are applied.

Bubble Sets
LineSets

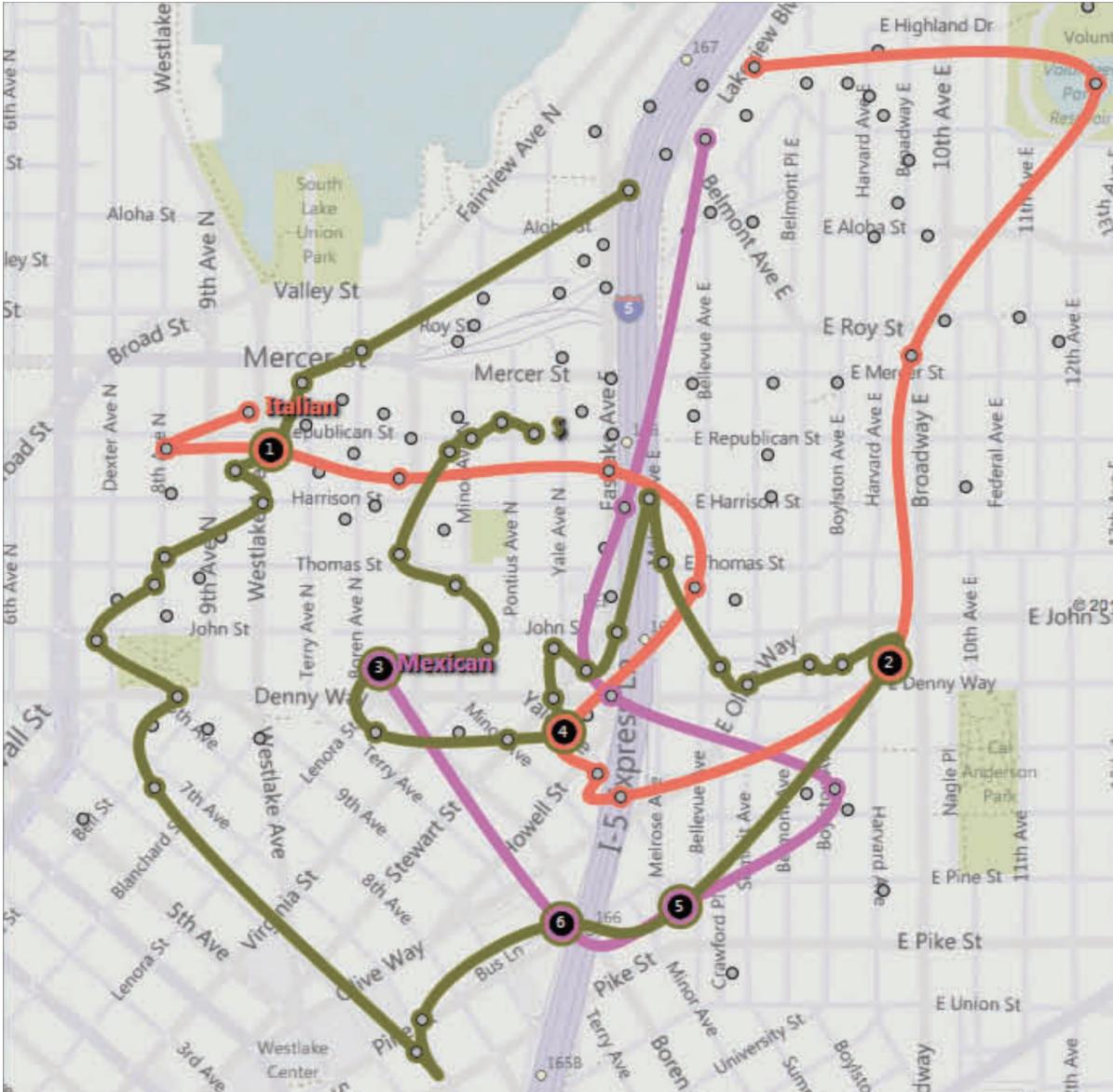
Bubble Sets

Collins, Christopher, et al. Bubble sets: Revealing set relations with isocontours over existing visualizations. IEEE Transactions on Visualization & Computer Graphics 6 (2009): 1009–1016.



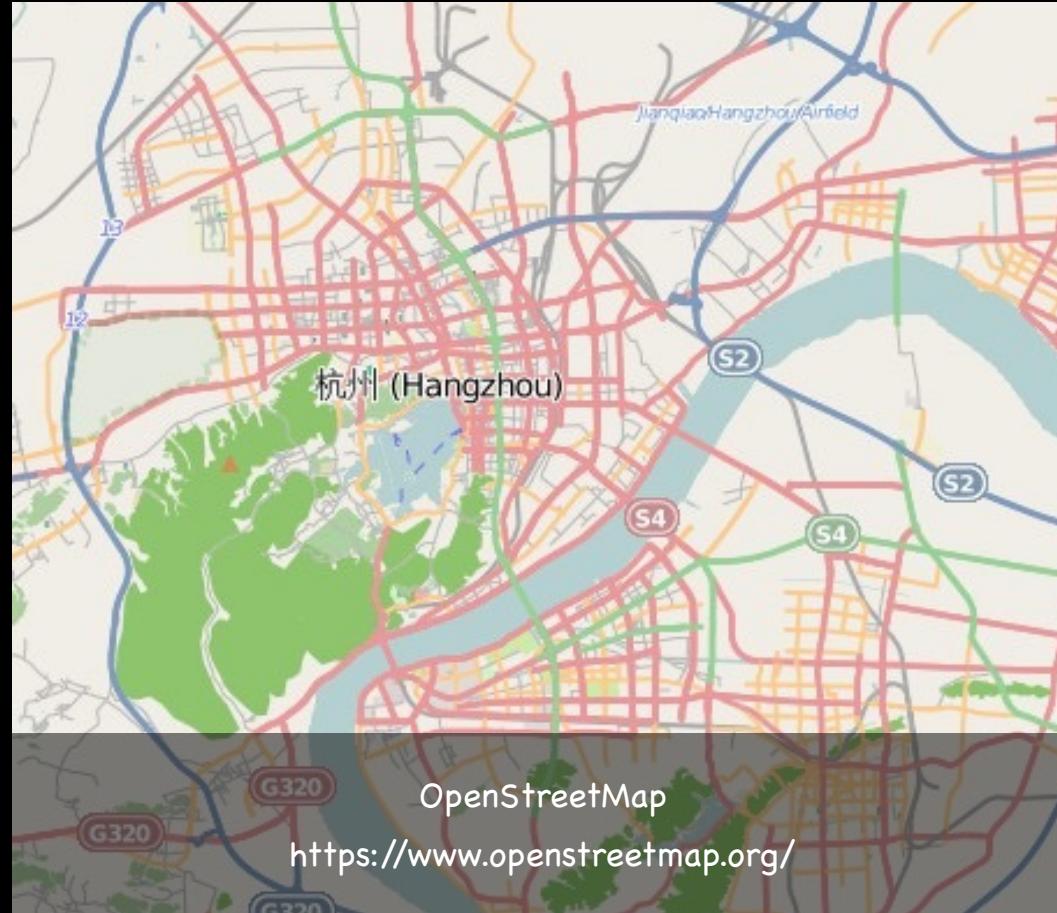
LineSets

Alper, Basak, et al. Design study of linesets, a novel set visualization technique. IEEE Transactions on Visualization & Computer Graphics 12 (2011): 2259-2267.

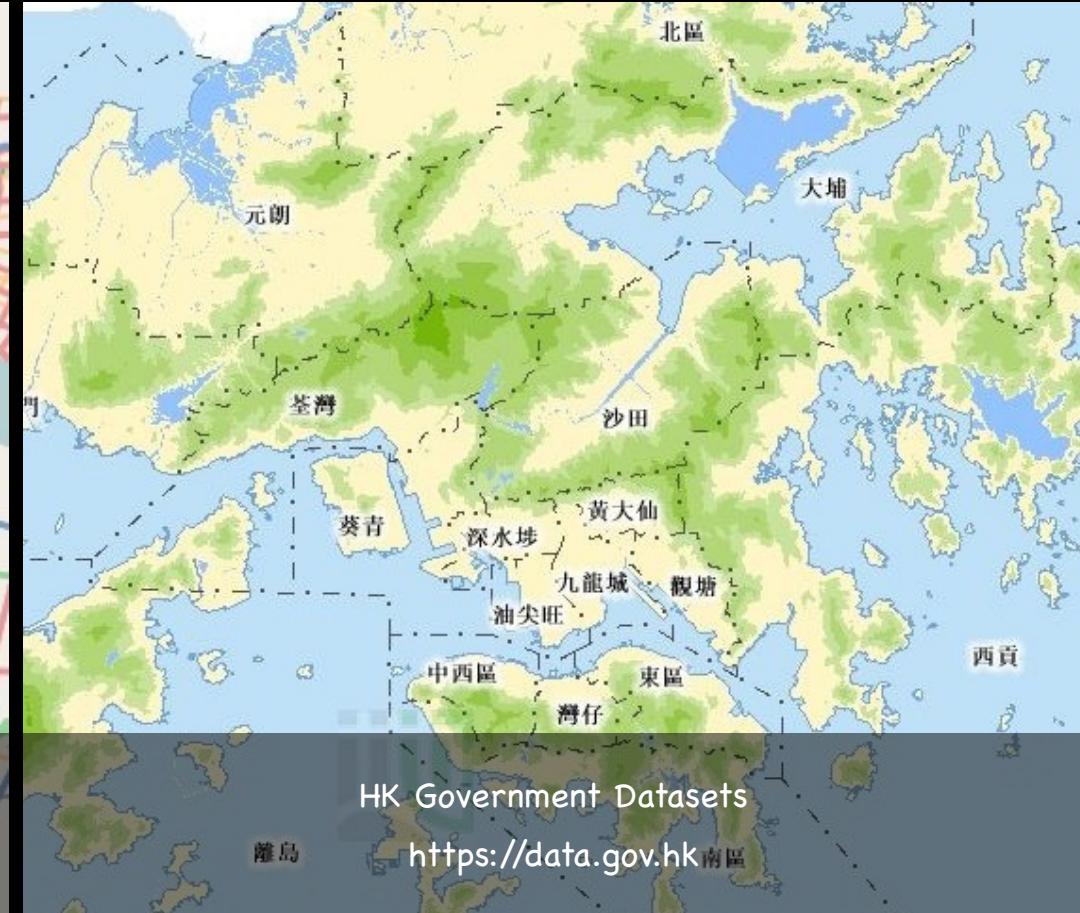


Geographic Datasets

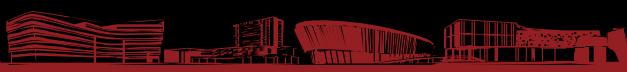
Geographic Datasets



OpenStreetMap
<https://www.openstreetmap.org/>



HK Government Datasets
<https://data.gov.hk>



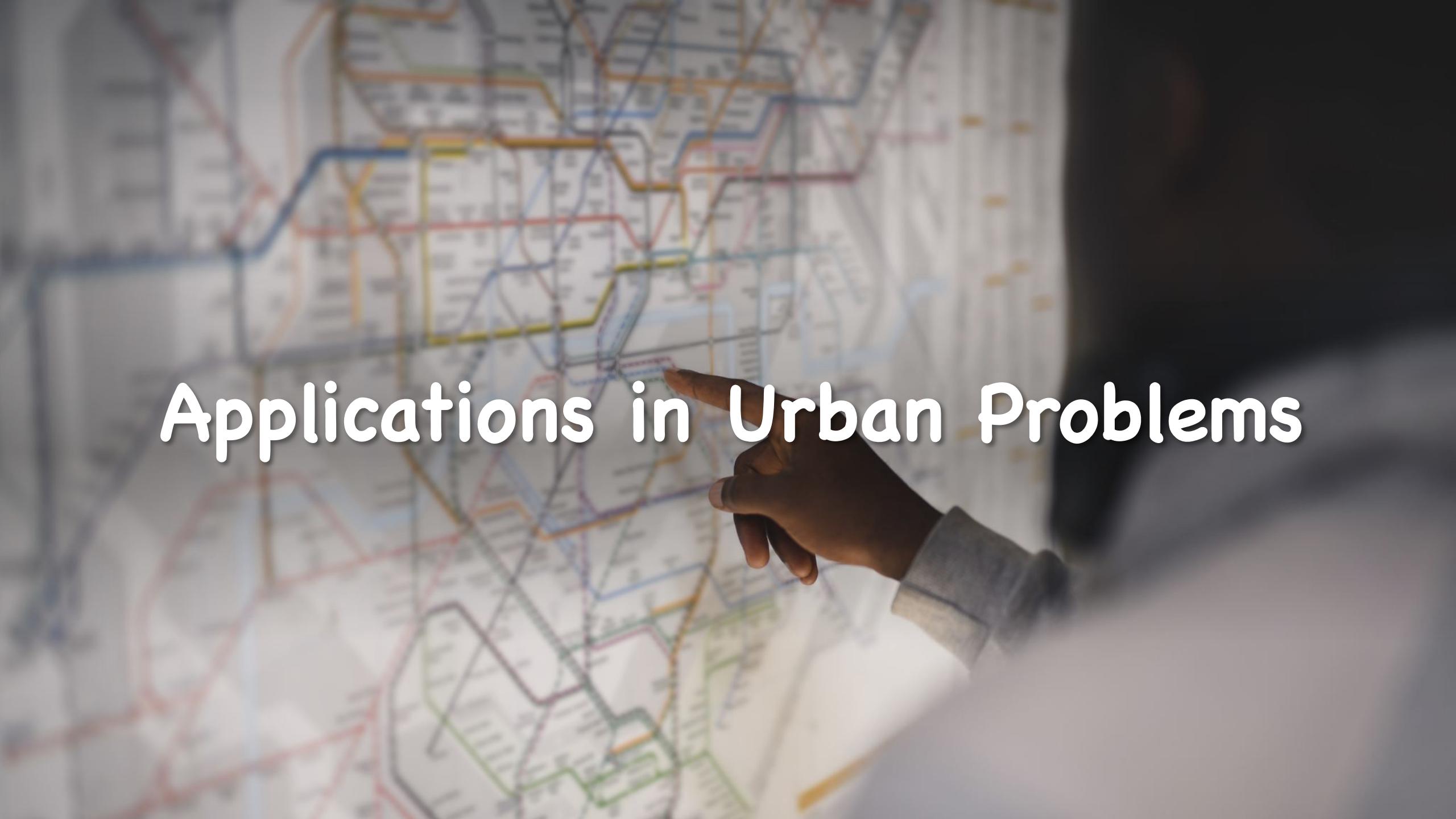


Geographic Datasets

- Google/Baidu API
- OpenStreetMap (<http://www.openstreetmap.org>)
- HK government dataset (<http://data.gov.hk>)
- NYC Open Data (<https://opendata.cityofnewyork.us>)
- Data.gov (<https://www.data.gov>)



Applications in Urban Problems





Big Challenges in Big Cities

Urbanization

- 1900 | 2 out of every 10 people lived in an urban area
- 1990 | 4 out of every 10 people lived in an urban area
- 2010 | 5 out of every 10 people lived in an urban area
- 2030 | 6 out of every 10 people will live in an urban area
- 2050 | 7 out of every 10 people will live in an urban area

Defined by UN HABITAT



Smart Cities – Big Urban Data

- Smart cities: utilize information technologies to **accurately analyze** and **efficiently solve** the challenges.



Heterogeneous

High-dimensional

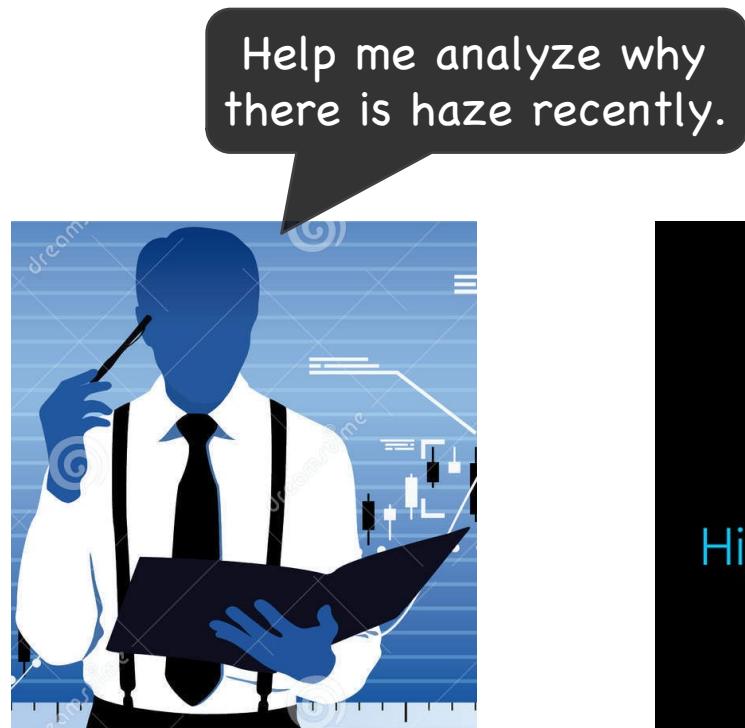
Dynamic, uncertain

Globally sparse

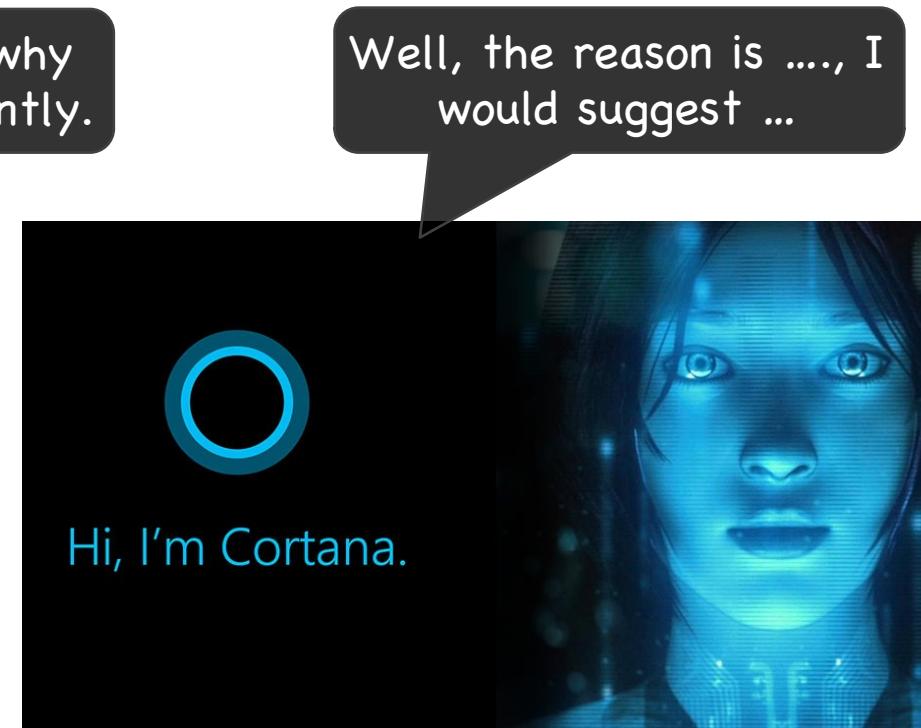
Locally redundant

Space-time break

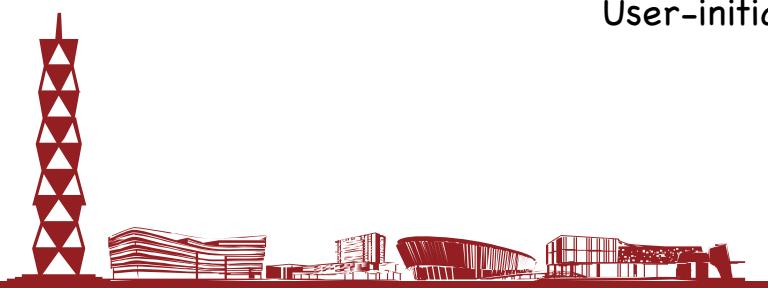
An “Ideal” Solution?



User-initiative



System-initiative





A 3D rendering of a person standing at the entrance of a dark, multi-level labyrinth. The person is a simple blue humanoid figure. The labyrinth is composed of dark grey and black walls with white highlights, creating a sense of depth and complexity. The floor is a light grey color.

Almost impossible in near future

Key Problems (1/3)

Data Source	Property	Value Range
Air Quality	PM 2.5, PM 10, O ₃ (ug/m ³)	[0, 500]
	NO ₂ (ug/m ³)	[0, 300]
	CO (ug/m ³)	[0, 20]
	SO ₂ (ug/m ³)	[0, 50]
	Level	6 levels
Weather	Temperature °C	[-20, 40]
	Humidity	[0, 100]
	Wind Speed (m/s)	[0, 300]
	Wind Direction	[1, 24]
	Cloud Conditions	14 conditions
Traffic	Total Cars	[9, 200]
	Low Speed (0 ~ 20 km/h) %	[6, 80]
	Medium Speed (20 ~ 50 km/h) %	[17, 74]
	High Speed (above 50 km/h) %	[0, 53]

- How to **visually summarize** large, complex data.
 - “Show me a visual summary”

Key Problems (2/3)

- How to leverage knowledge to derive new insights.
 - “It is not the answer I expect. I want to... do it again”



Too many trajectories. I want to remove ...

This assumption is too strong. I want to ...

These data points are irrelevant. I want to ...

The result is not right. I want to try...

This pattern is suspicious. I want to ...



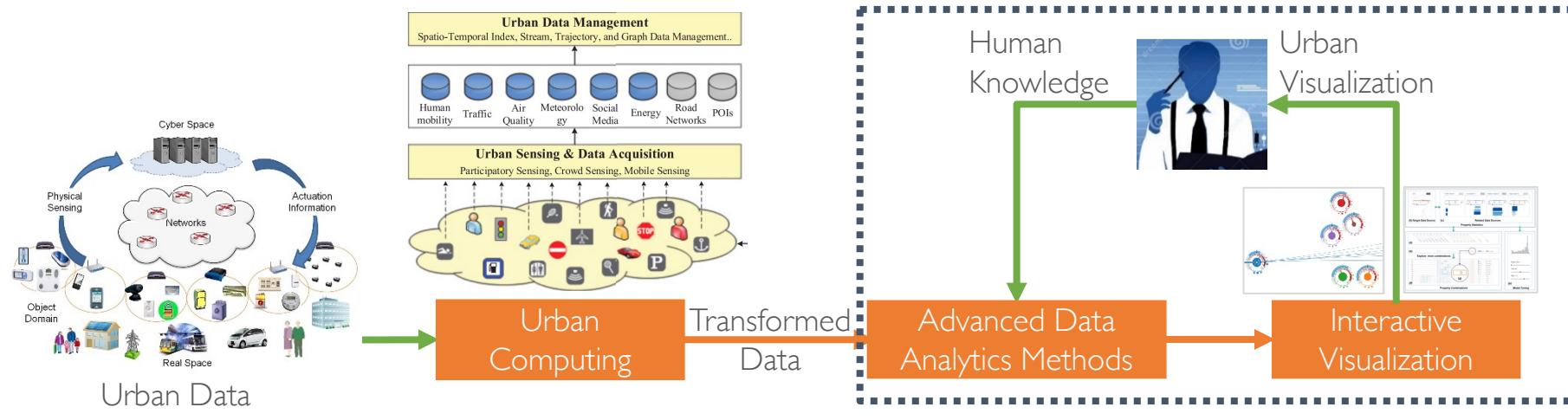
Key Problems (3/3)

- How to convey and explain the discovered insights.
 - “Interesting insights. Tell me why...”



Visual Analytics of Urban Data

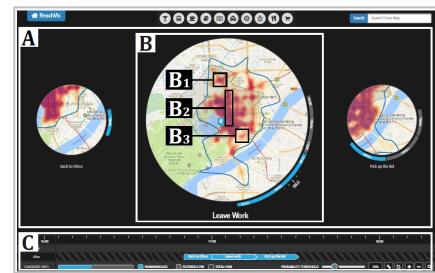
- **Visual urban analytics** combining advanced data analytics methods and interactive visualization.
 - Fast, easy, multi-level visualization
 - Rich analyses by interactive visualization
 - Storytelling experience



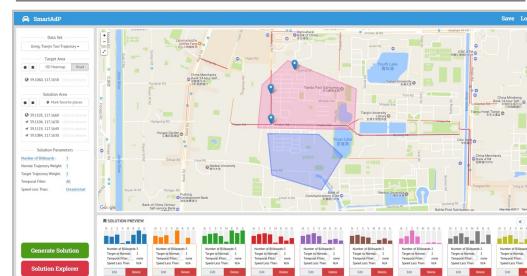
Urban Planning



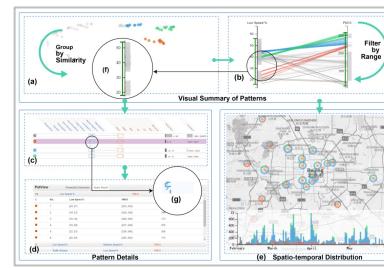
Safety and Security



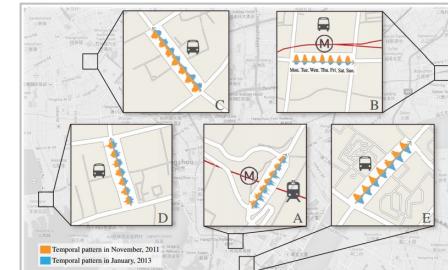
Business



Environment



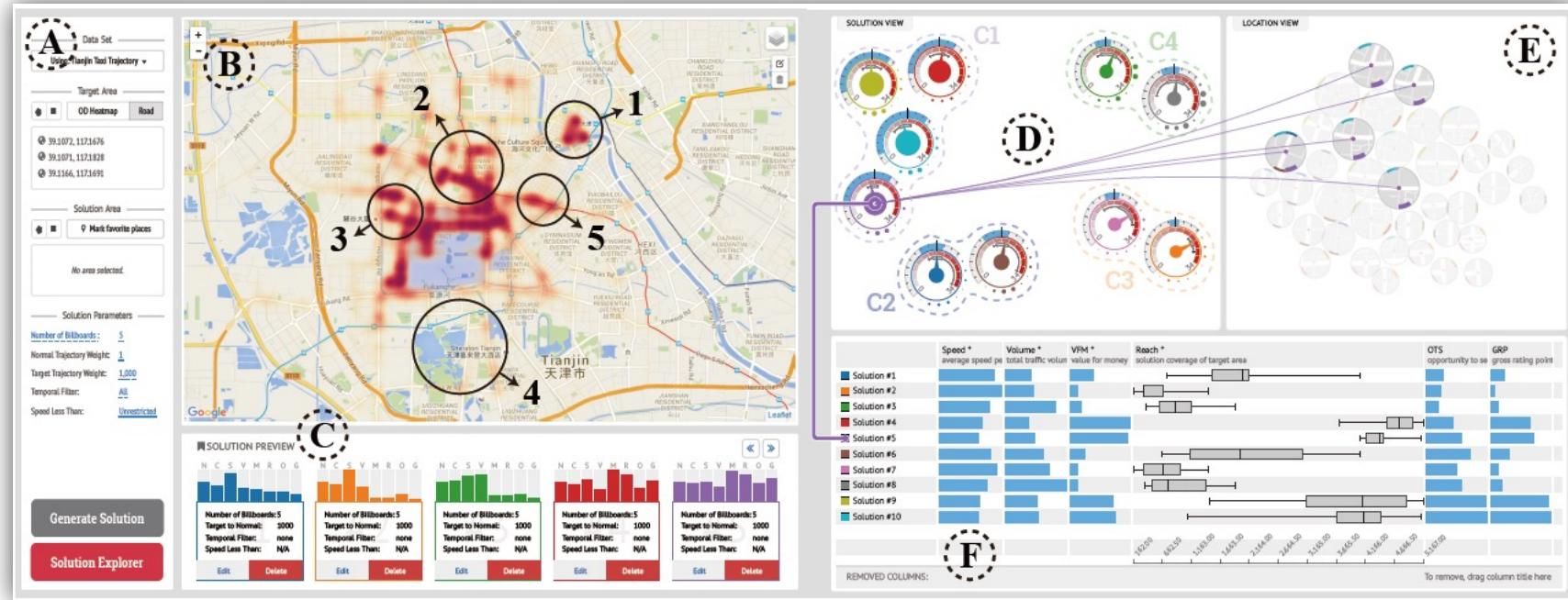
Transportation



Business - Location Selection



上海科技大学
ShanghaiTech University



SmartAdP: Visual Analytics of Large-scale Taxi Trajectories
for Selecting Billboard Locations

IEEE VAST 2016 (IEEE TVCG)



Microsoft®
Research
微软亚洲研究院

立志成才报国裕民

Billboard Advertising

- Billboards work because it is seen by everyone on the roads.



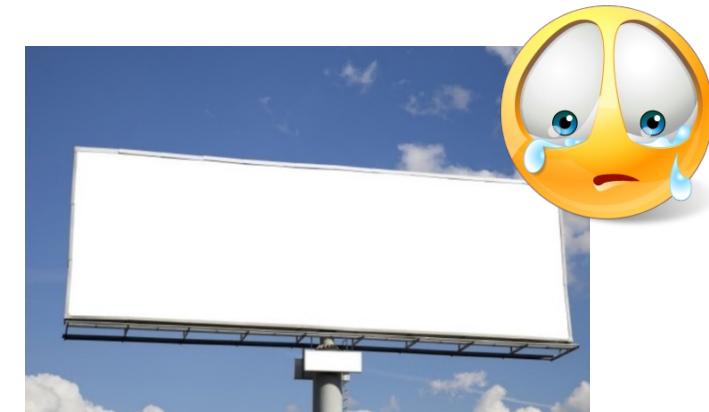
Billboard Location



Location,
Location,
Location



Appropriate locations



Inappropriate locations



立志成才报国裕民

Factors To Choose A Right Location

- What factors to consider?

Billboard Location Selection Factors



上海科技大学
ShanghaiTech University

Traffic volume

Traffic speed

Traffic OD

Environment



立志成才报国裕民

Billboard Location Selection Factors



上海科技大学
ShanghaiTech University

Traffic volume

Traffic speed

Traffic OD

Environment



立志成才报国裕民

Billboard Location Selection Factors

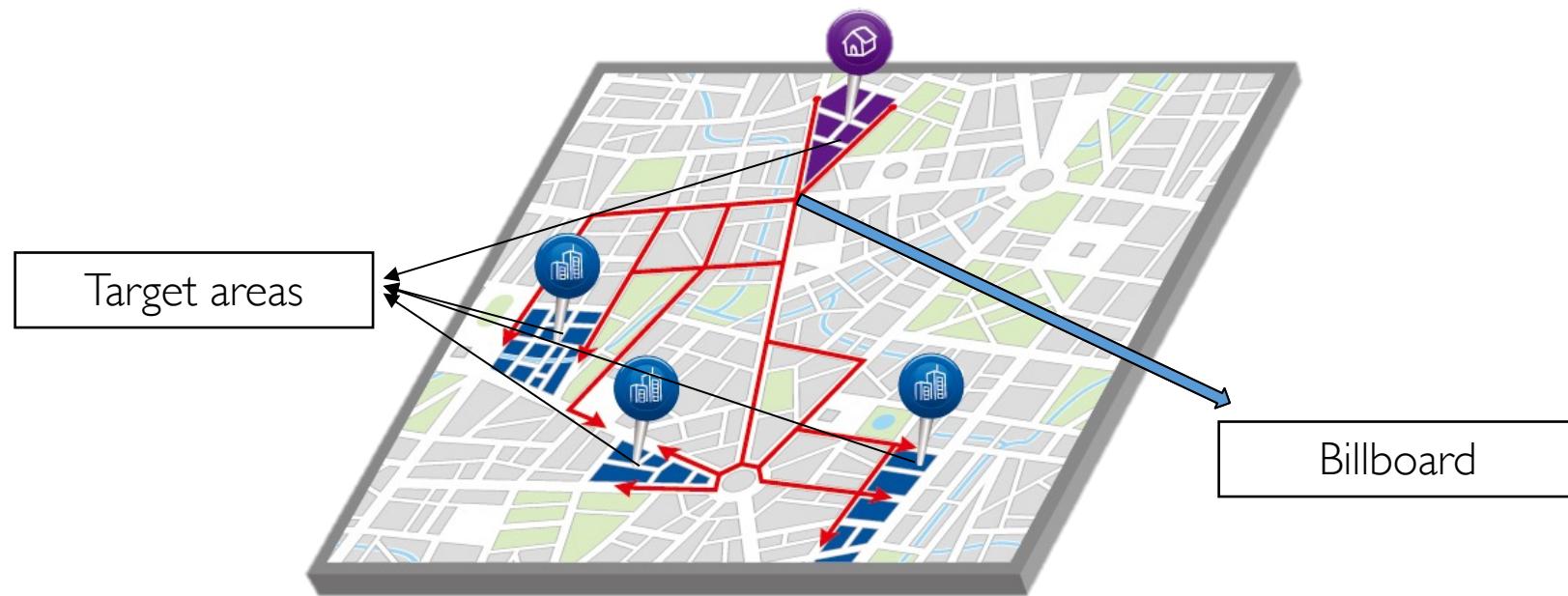


Traffic volume

Traffic speed

Traffic OD

Environment





Billboard Location Selection Factors

Traffic volume

Traffic speed

Traffic OD

Environment



Home



Work



Shopping
& Errands



Dining &
Entertainment



Education
Health &
Professional
Services



Key Challenges (1/3)

- How to acquire such comprehensive data cheaply.
 - Conventional approach is expensive and time-consuming.



Travel Surveys:

- 1) Non-Work Destination Type Data
- 2) Non-Work Trip Frequency Data
- 3) Used to model non-work trip destinations and frequencies

U.S. Census:

- 1) Demographic Data
- 2) Work Destinations
- 3) Work Trip Frequency Data
- 4) All data from individuals in specific block groups

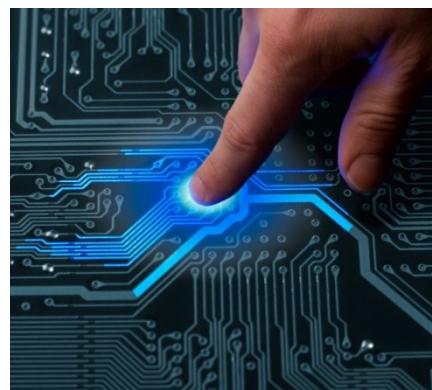


Easy to collect

Reveal underlying traffic patterns

Key Challenges (2/3)

- How to formulate a good solution quickly?
 - Large solution space
 - Some factors are hard to model
 - Different opinions on the optimal



Computational Power

+



Human Knowledge

=

Cost-effective
Advertising Plan



立志成才报国裕民

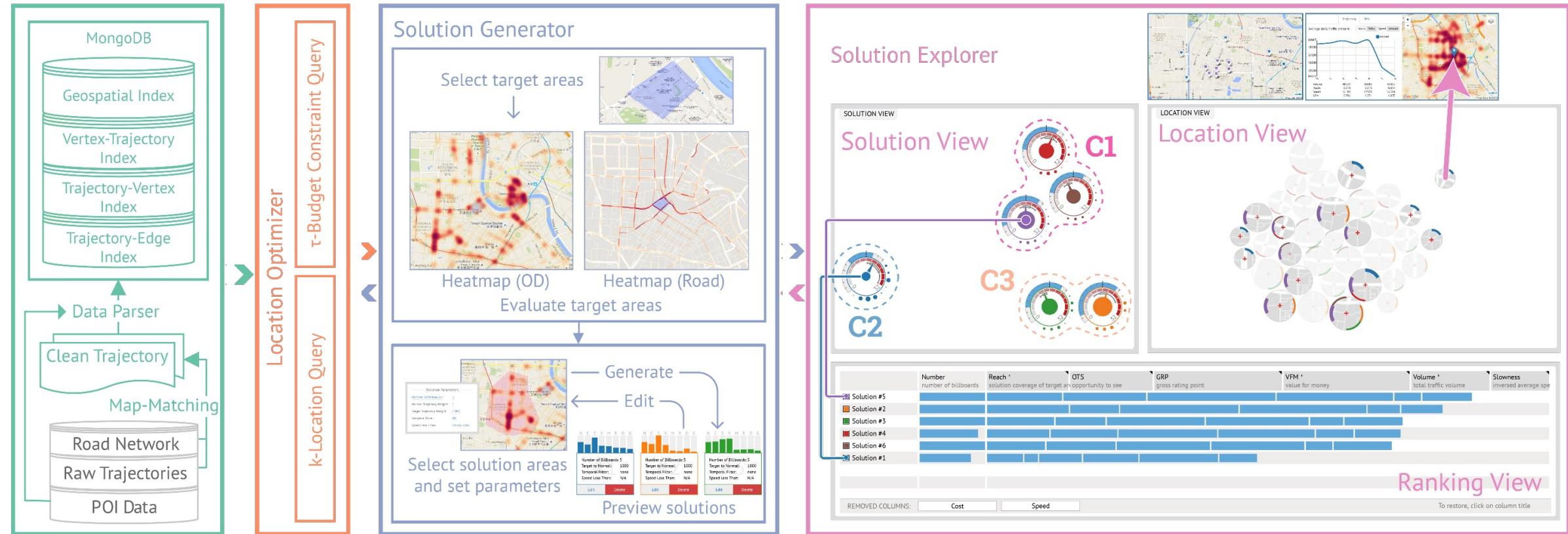
Key Challenges (3/3)

- How to visually compare several candidate solutions to find the optimal one.
 - Which advertising strategy is better...
 - What if I spend more \$50,000...



Our Solution

- Visual analytics system combining advanced data mining methods and interactive visualization techniques.





Analytical Tasks

Location Optimizer

Solution Generator

Solution Explorer

Which areas befitting for placing billboards?

Where are the optimal locations?

How good is a billboard location/solution?

What are the differences and similarities?

How many groups exists?

What is the rank? Which ones are optimal?



Solution Generator

Data Structure Construction



- To extract information from large scale trajectories and speed up the mining algorithm, we build three indexes.

Edge (road segment)		
Trajectory	Tr _i	
	e_{i1}	e_{i2}
		\dots

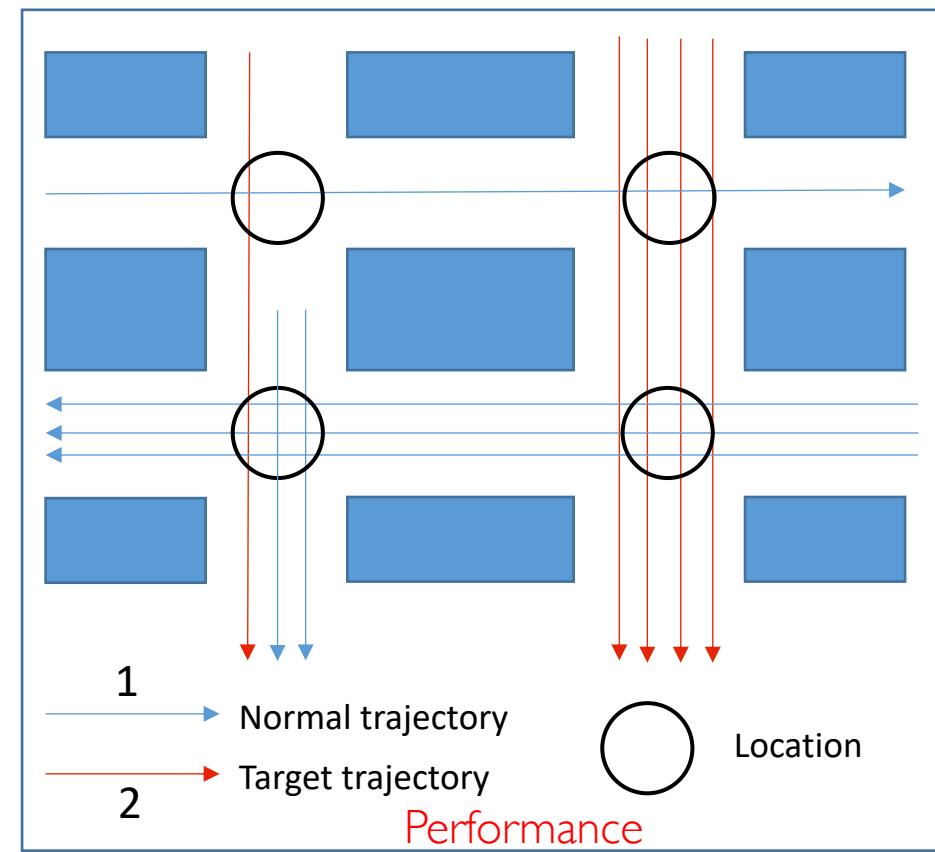
Potential billboard location Vertex (road intersection)		
Trajectory	Tr _i	
	v_{i1}	v_{i2}
		\dots

Trajectory		
Trajectory	V _i	
	Tr_{i1}	Tr_{i2}
		\dots



Mining Algorithm

- Problem: identify a set of locations with the maximum trajectory coverage (the sum of the weights of the covered trajectories)
 - Each trajectory has a weight
 - Each location has a cost
- K-Location Query
 - Selecting & updating
- τ -Budget Constraint Query
 - Utilization ratio:
 - $U(v) = \frac{\text{current coverage value}}{\text{cost}}$



guarantee: $1 - 1/e$

立志成才报国裕民

Analytical Tasks

Location
Optimizer

Solution
Generator

Which areas befitting for placing billboards?
Where are the optimal locations?
How good is a billboard location/solution?

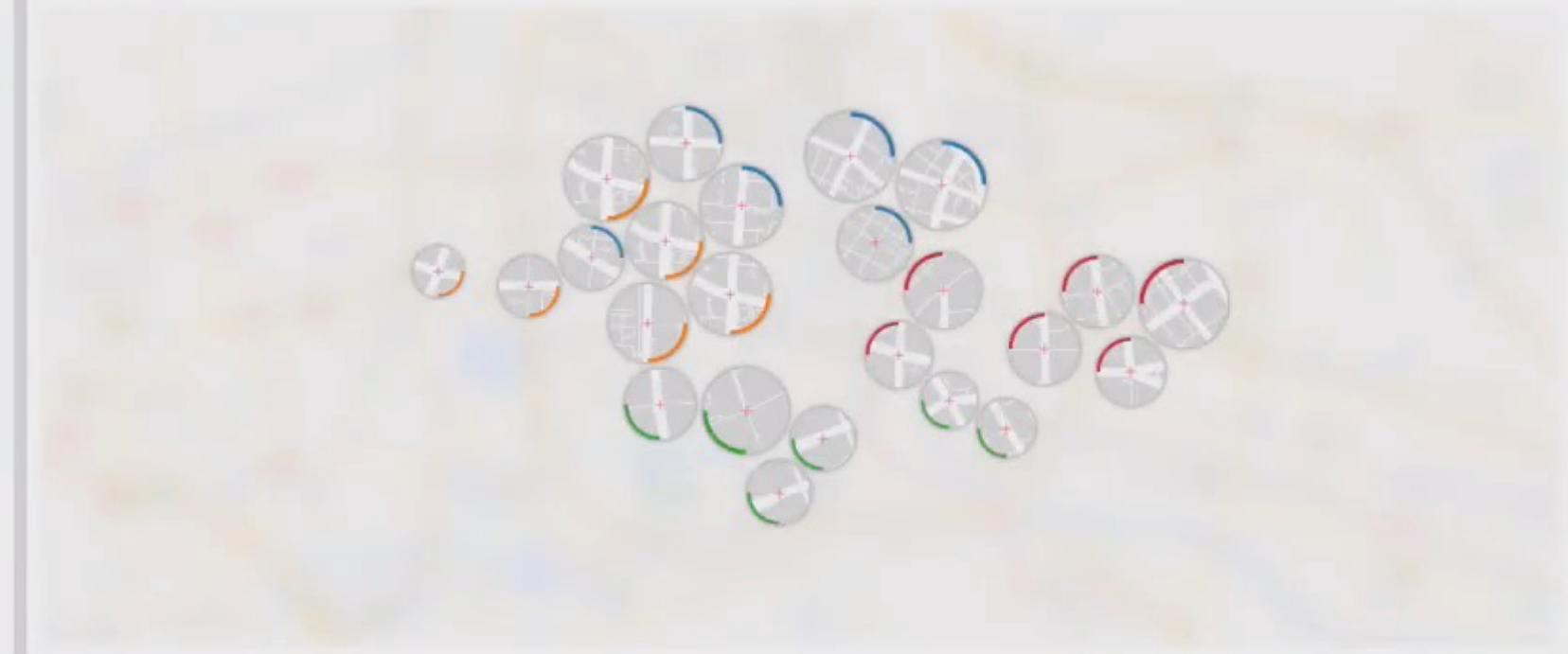
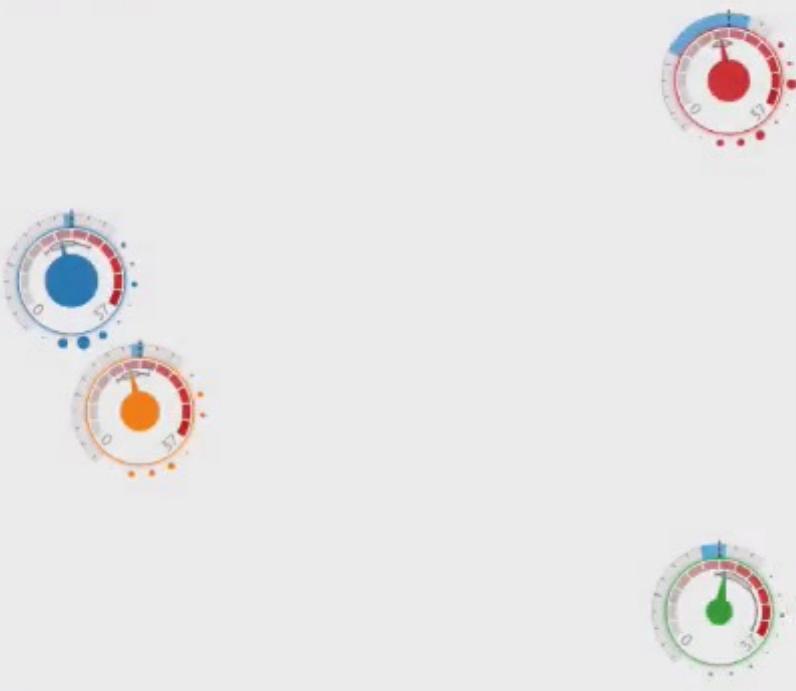
Solution
Explorer

What are the differences and similarities?
How many groups exists?
What is the rank? Which ones are optimal?



SOLUTION VIEW

EDICATION VIEW



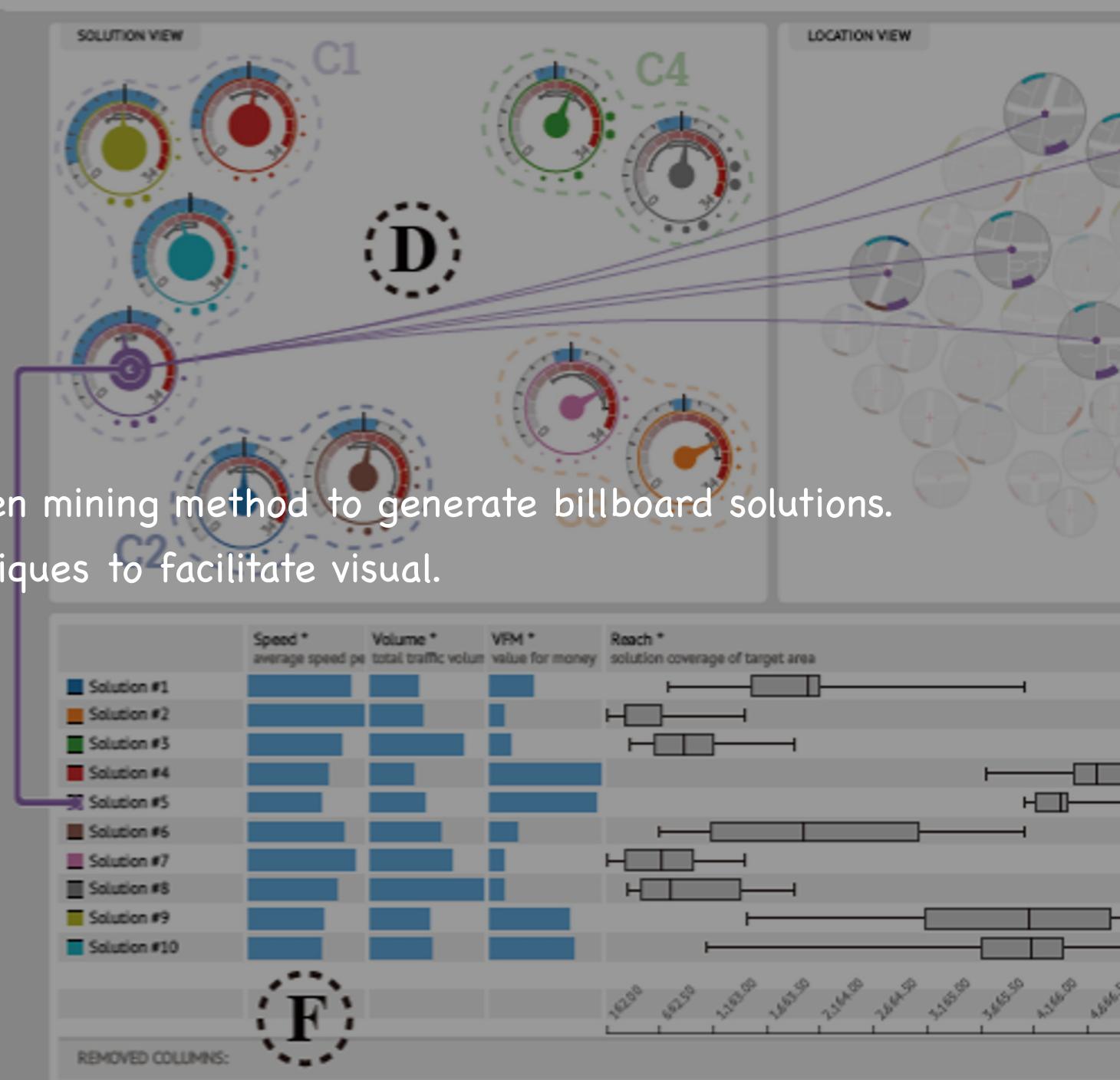
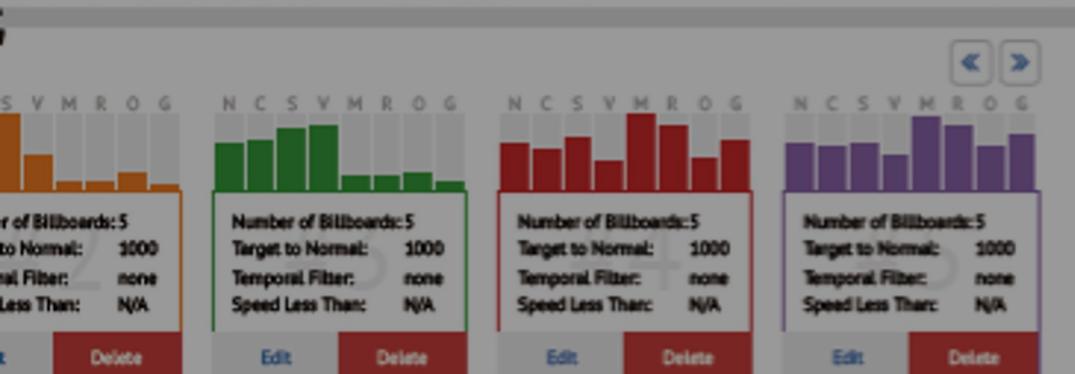
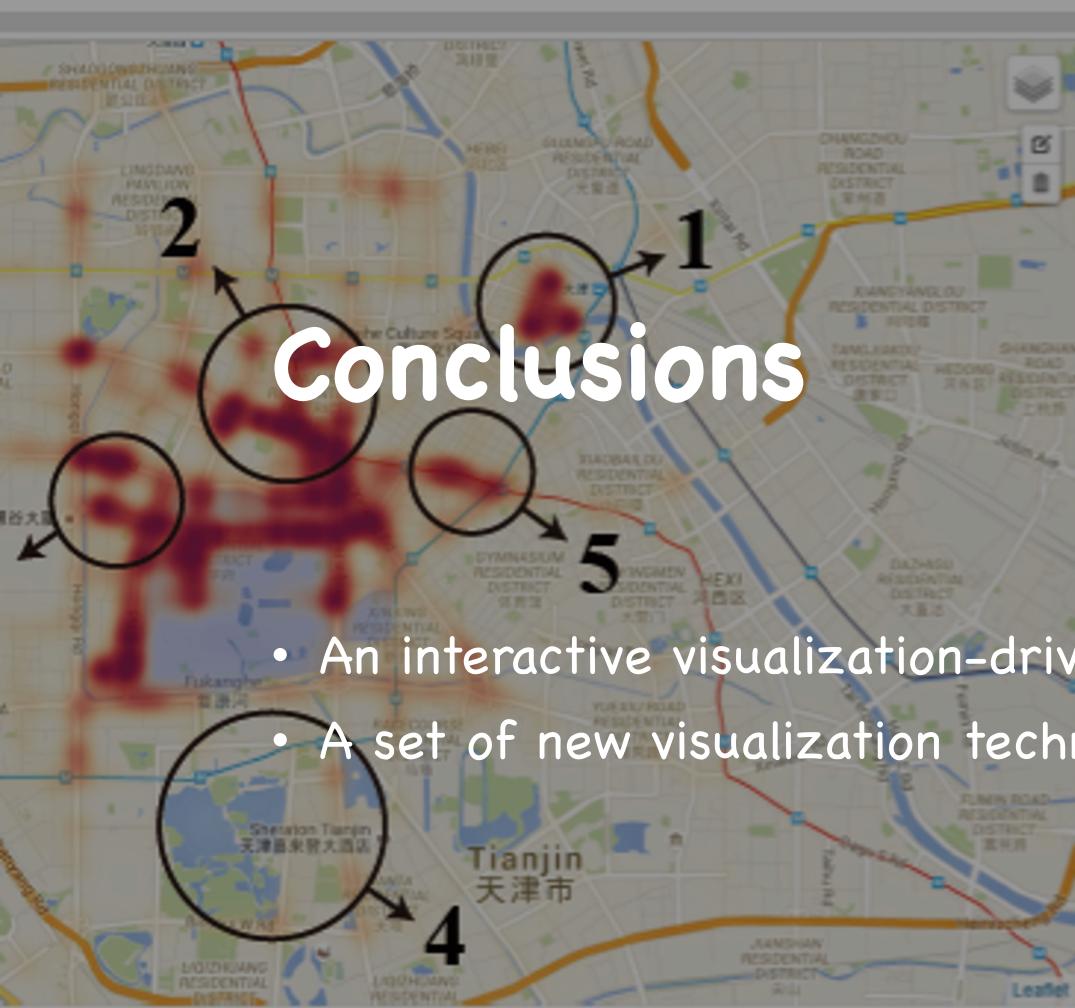
REMOVED COLUMNS:

To remove, drag column title here

Case Study

- Case study of the Tianjin dataset
 - Road network data, 133,726 road segments and 99,007 vertices
 - GPS trajectory data, 4 million trajectories
 - POI data, 154,633 points

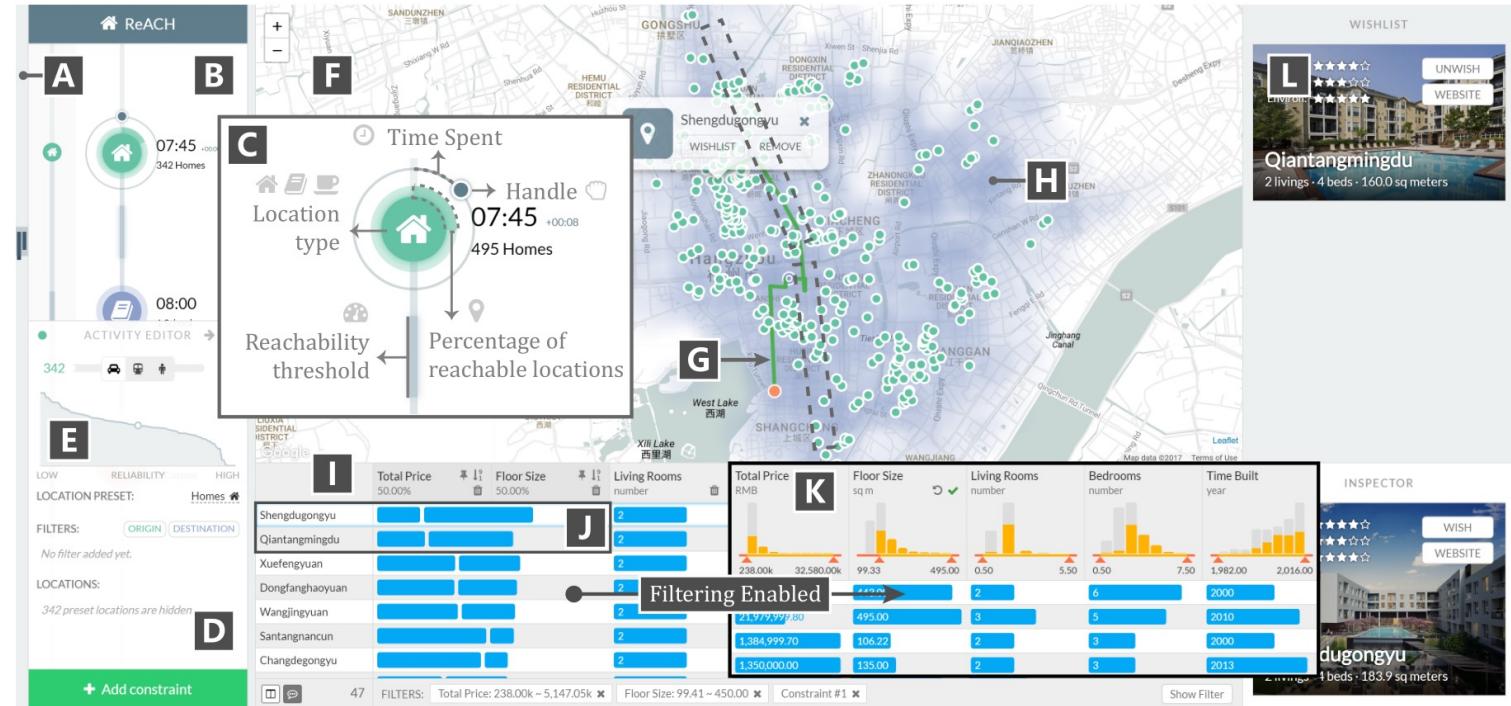
Case Studies



Business - HomeFinder



上海科技大学
ShanghaiTech University



HomeFinder Revisited: Finding Ideal Homes with Reachability-Centric Multi-Criteria Decision Making

ACM SIGCHI 2018



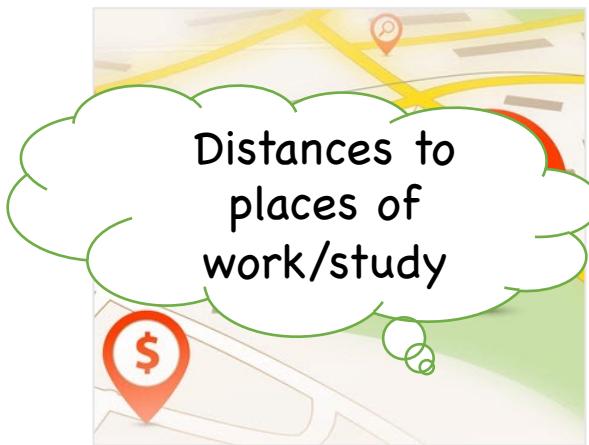
Microsoft Research
微软亚洲研究院

立志成才报国裕民

Finding An Ideal Home...



- is a difficult and laborious process.



Location



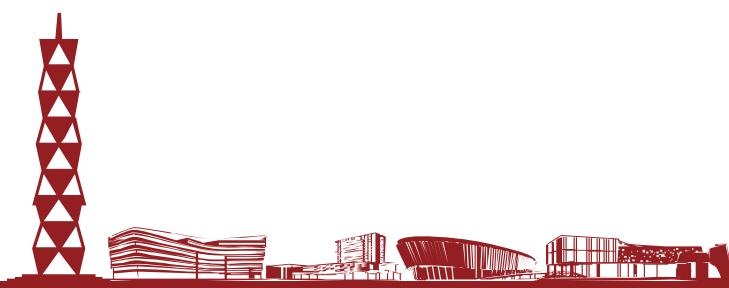
Price



Size & Layout

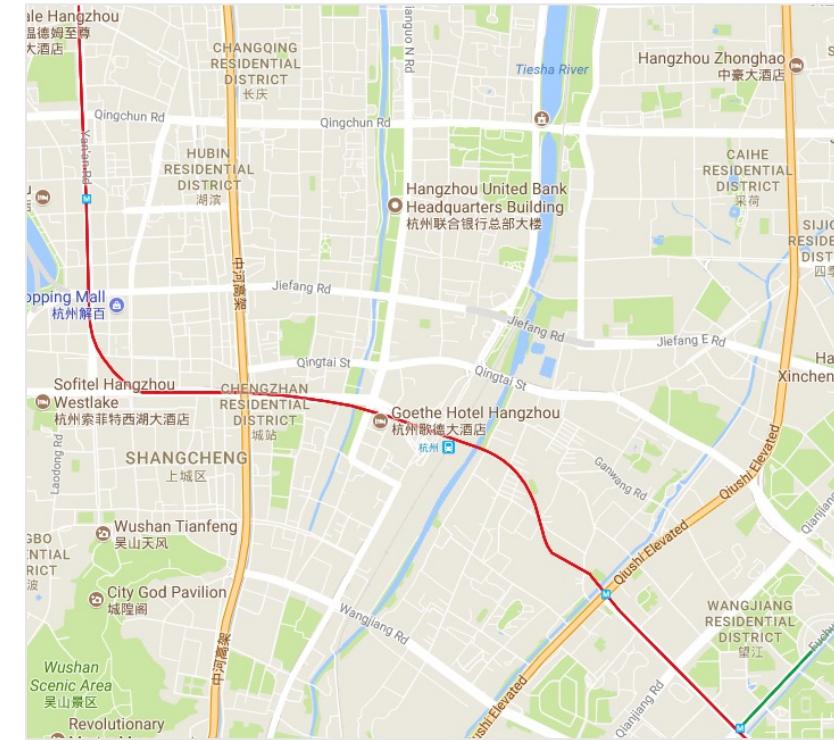
And much much more...

Location is one of the foremost factors involved.



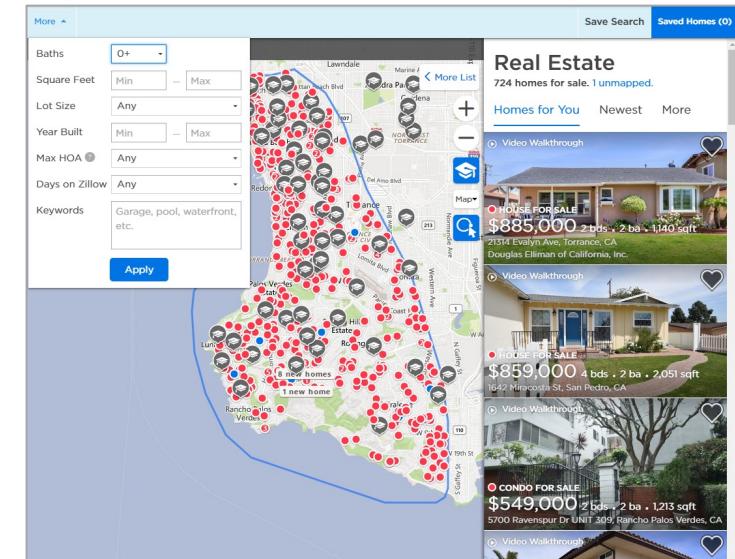
Location, Location, Location

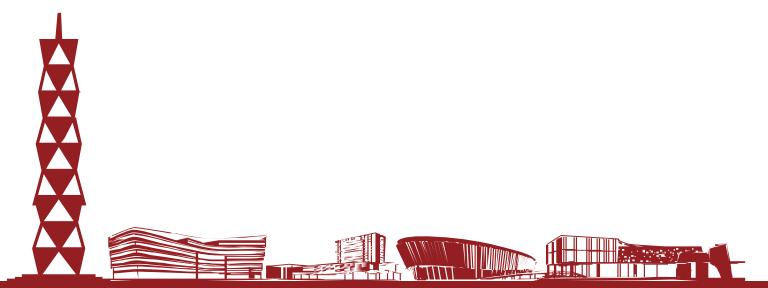
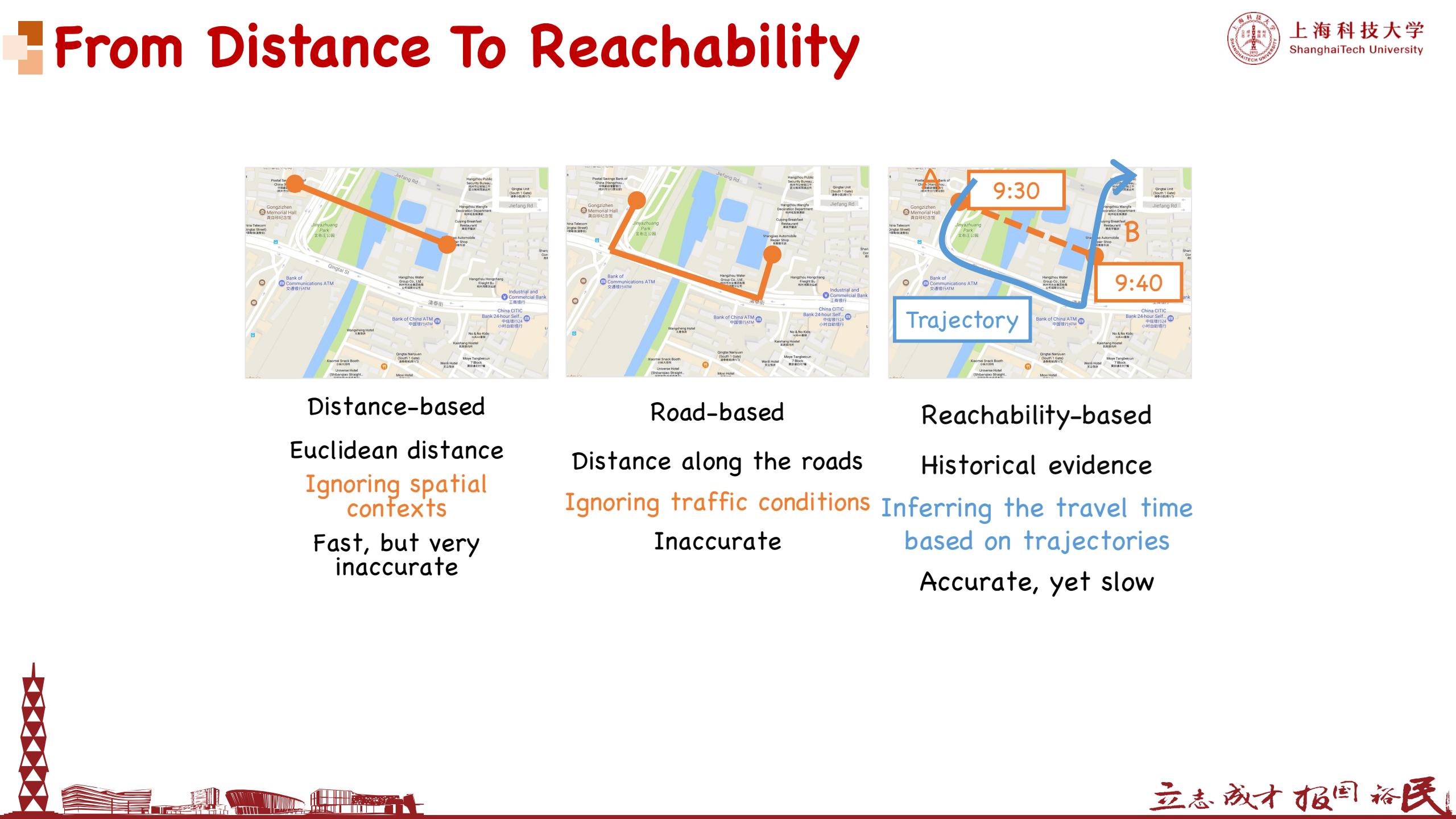
- Find me a house so I can...
 - Leave home at **8:15 a.m.**
 - Drive the kid to school before **8:30 a.m.**
 - Arrive at work before **9:00 a.m.**



Revisiting HomeFinders

- Domestic: Lianjia, Sofun; Foreign: Zillow, Hubzu
 - × Simple spatial search functions
 - × Spatial filtering are based on *distance*







	Total Price RMB	Floor Size sq m	Living Rooms number	Bedrooms number	Time Built year
	238.00k - 32,580.00k	18.50 - 495.00	0.50 - 5.50	0.50 - 7.50	1,982.00 - 2,016.00
江南水乡	14,000,000.53	443.00	2	6	2000
江南水乡	21,979,999.80	495.00	3	5	2010
荷韵江南苑	1,350,000.35	95.00	2	3	2009
绿都御景蓝湾	572,688.96	70.99	2	2	2014
绿都御景蓝湾	905,454.55	84.00	2	3	2015
绿都御景蓝湾	1,361,999.66	97.20	2	4	2012



1927

FILTERS:



Key Challenges

- Problem:
 - Identify ideal homes based on multiple criteria
 - including reachability that satisfies users' daily routines.

Efficiency of
reachability computation

"I don't have all day to see my mouse spinning."

Representation of
daily routines

"I would like to visit B before A. Determine a B for me."

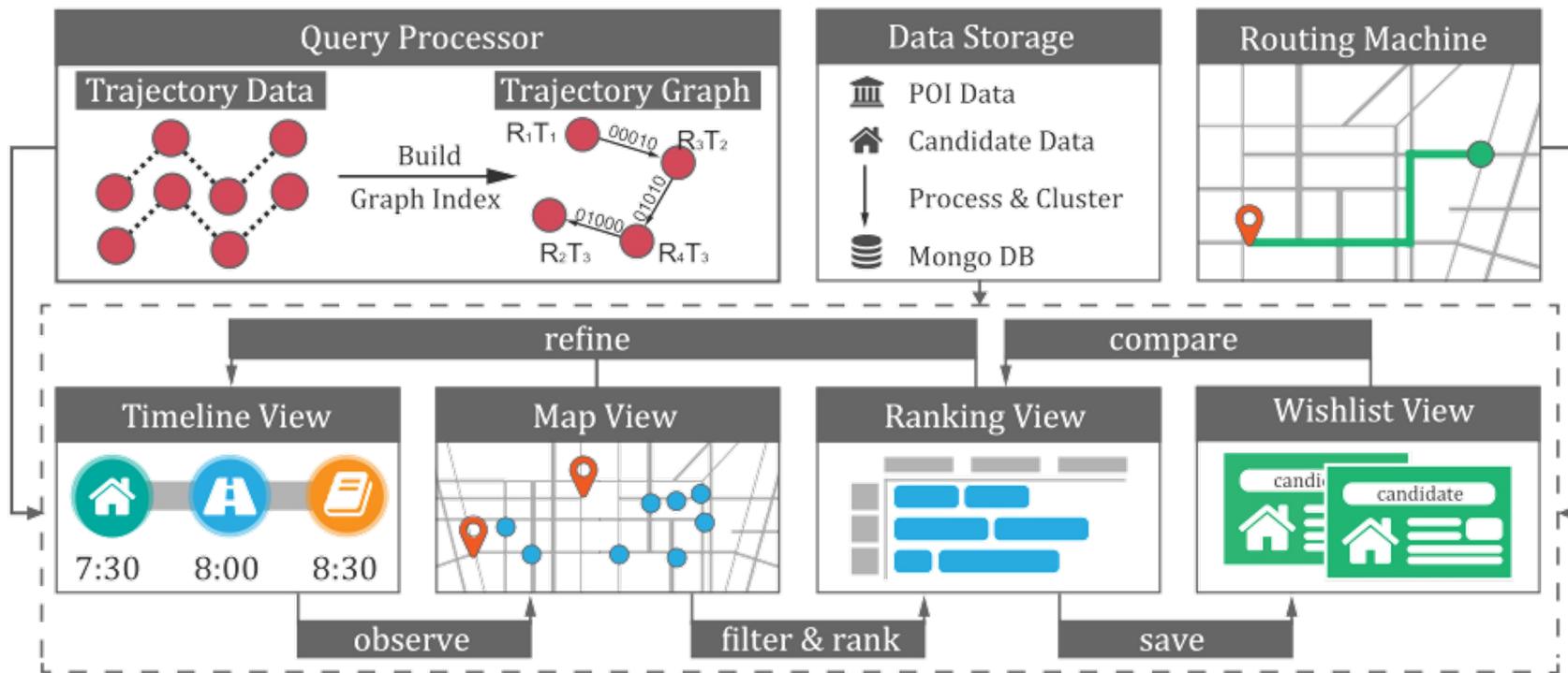
Integration of
individual preferences

"I want more bedroom. Can you find less pricy ones?"



Our Approach

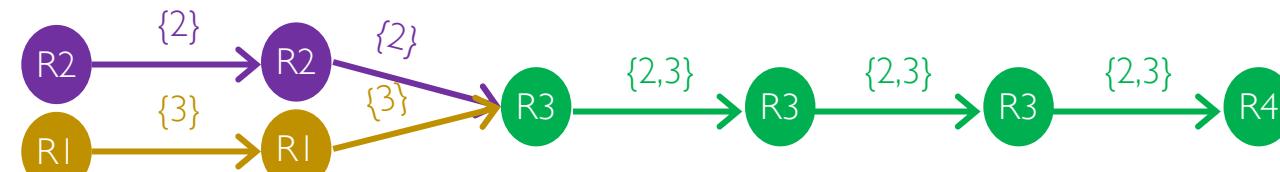
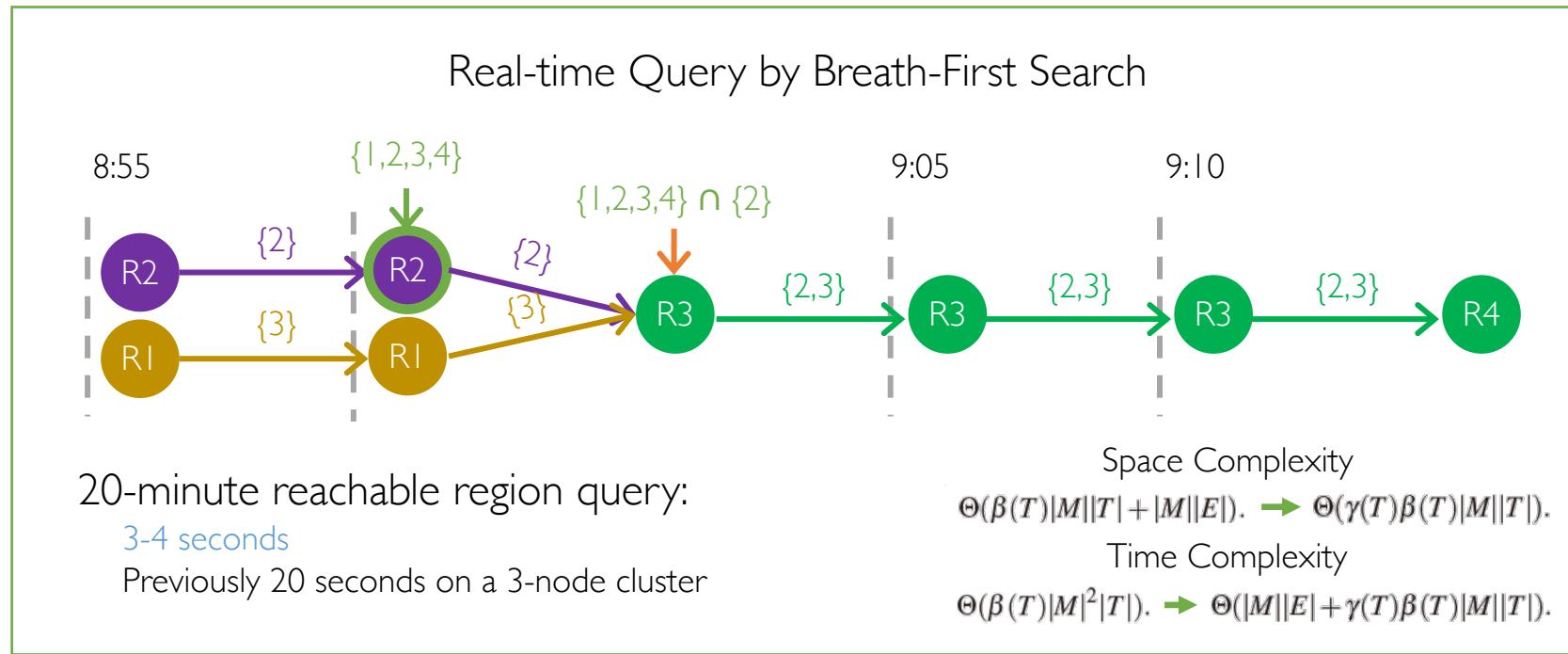
- Help average users **filter and analyze** candidate houses based on **the reachability generated** by the mining model.





Computing Reachability on Graphs

- On which days I can reach road A from B between time slots i and j?





Visual Design



The figure displays four screenshots of the ReACH app interface:

- Timeline View:** Shows a timeline from 07:45 to 08:00 with a location marker and a count of 342 homes.
- Map View:** A map of Shanghai showing various residential districts and landmarks like West Lake and the Bund. It highlights the "Shengdugongyu" area with a blue circle and a callout for "Map View".
- Ranking View:** A table ranking properties by total price, showing details for Shengdugongyu, Jiantangmingdu, Jiefengyuan, Longfanghaoyuan, Jangjingyuan, Jintangncun, and Hangdegongyu.
- Wishlist View:** A card for "Qiantangmingdu" listing 2 living rooms, 4 beds, and 160.0 sq meters, along with nearby, traffic, and environment ratings.

Timeline View

Map View

Wishlist View

Ranking View

Refine

Timeline
View

Map View

Ranking
View

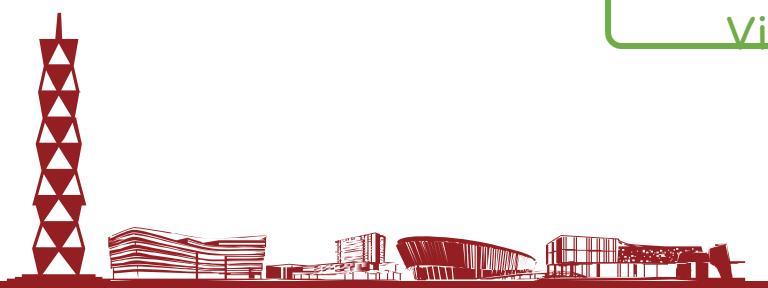
Compare

Wishlist
View

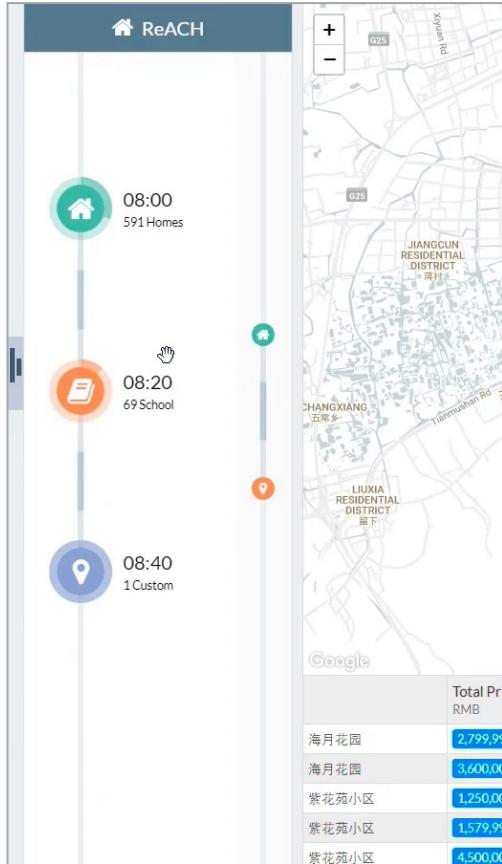
Observe

Filter & Rank

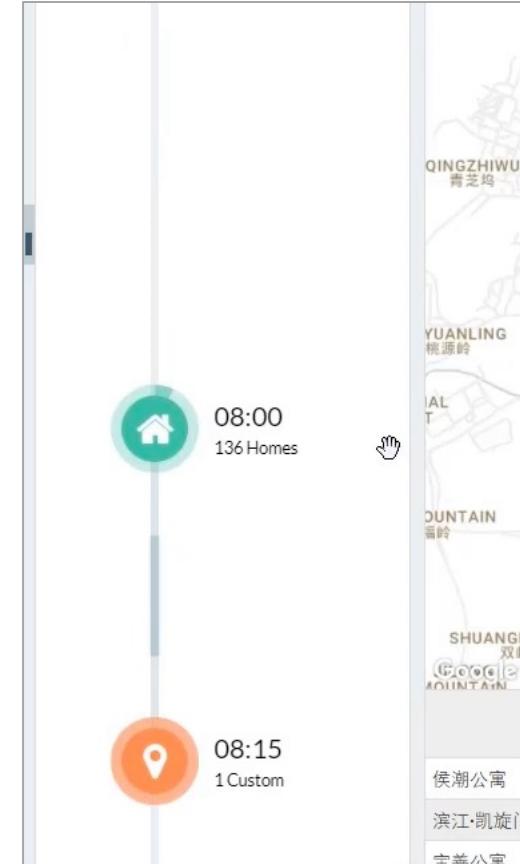
Save



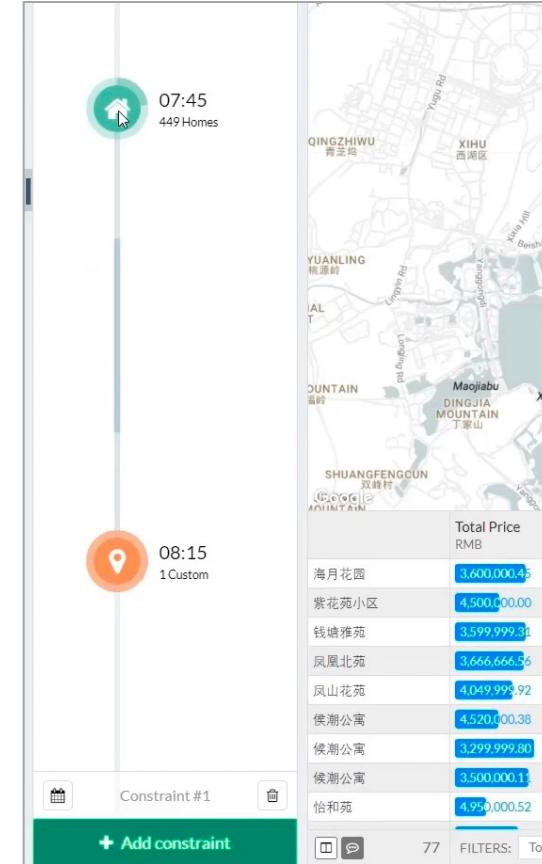
Interactions



Adaptive Zooming



Intuitive Dragging



Activity Editing



立志成才报国裕民

Earthquake Evacuation

- Tracking human mobility dynamics in a large-scale earthquake evacuation is of significance



- Understanding the evacuation behaviors remains limited due to the heterogeneous, spatiotemporal and multivariate urban data

Challenges

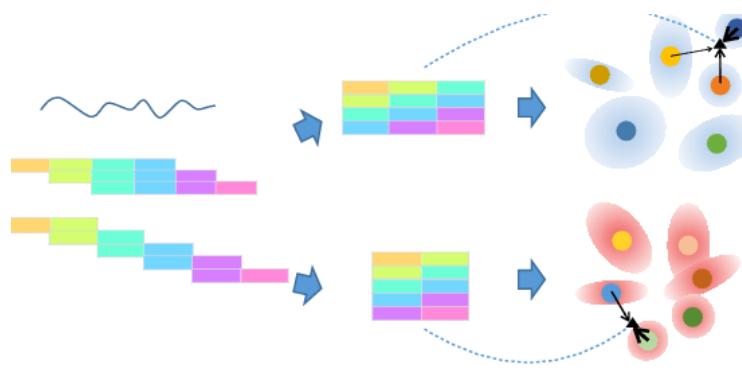
Massive Information

- Surrounding environment
 - Earthquake damages
- Large-scale human movements
 - Limited screen space



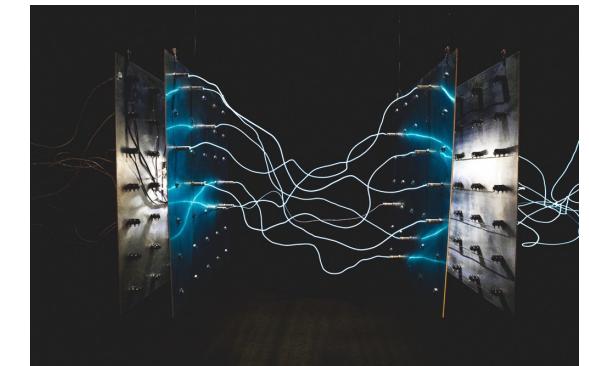
Feature Dynamics

- Multi-source and heterogeneous
 - Fuse and evolve
- Track and visualize dynamics in a 3D space



Feature Correlation

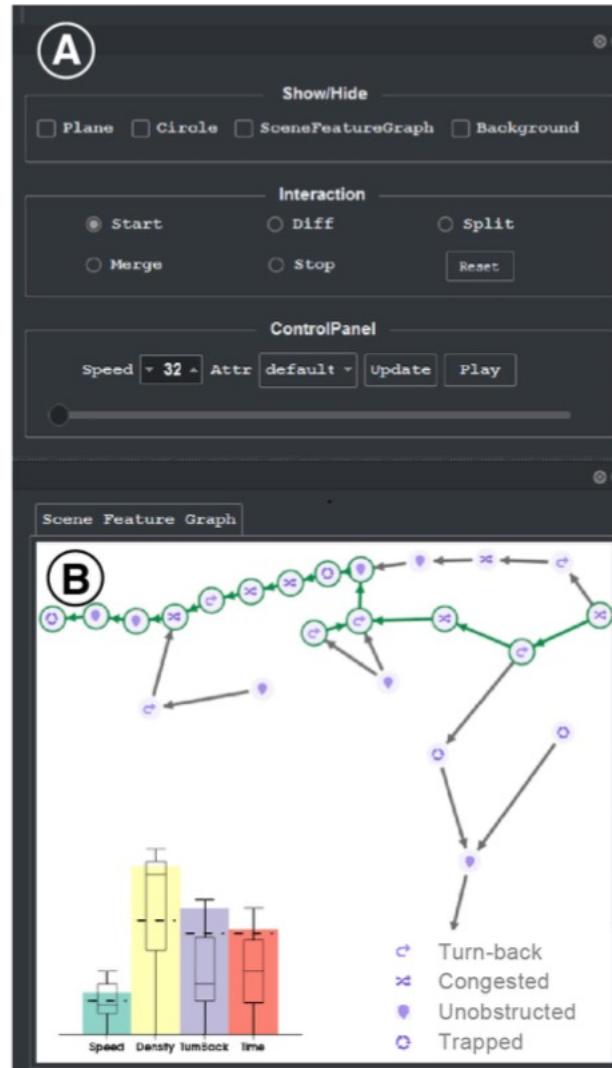
- Display surroundings
- Comparison among multiple simulations for a specific event
 - Examining correlation of features with surroundings



Overview of SEEVis

Storyboard View

Control Panel



Interactive Scene Feature Graph



Dataset Description and Term Definition

ENVIRONMENT

- A hybrid model of a high-resolution 2D grid
- A topological graph of the open space

AGENT

- Residents, visitors, officials, and cars that can see, think and act are all defined as agents

FEATURES

- Various implicit attributes such as speed, velocity deviation, density, evacuation time, and its variation

EVENTS

- Turn-back
- Congestion
- Unobstructed
- Trapped

Purpose Acquisition Module

Step 1: Trajectory extraction

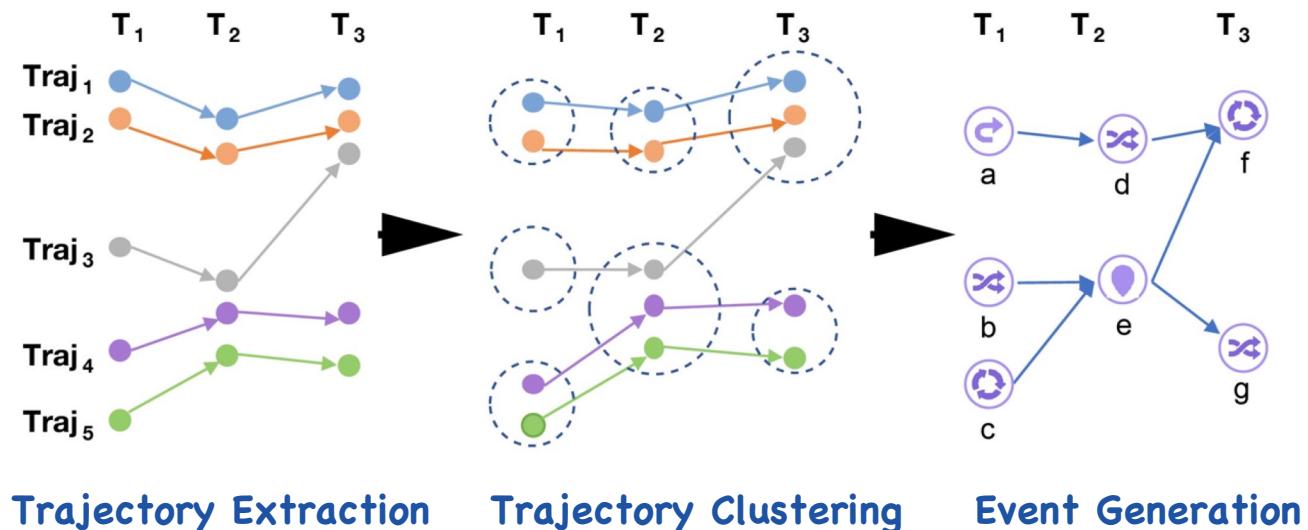
- Click an area of interest on the map
- Trajectories relevant to this area are extracted
- Underlying data features are pre-calculated

Step 2: Trajectory clustering

- Spatiotemporal clustering on the trajectories to generate a scene feature graph

Step 3: Events generation

- Feature event mapping to obtain the most dominating feature for each evacuation group
- Use the most dominating feature to represent the event for group



$$\text{icon}(v) = \operatorname{argmax}_i \frac{\text{feature}_i(v) - \text{threshold}_i}{\text{threshold}_i}$$



Trapped



Turn-back



Congested



Unobstructed

Storyboard Generation Module

- Input: scene feature graph
- Input: users' preferred features or regions
- Output: optimal path

Interaction Mechanism 1

- Indicate preferred feature
- Calculate path with the largest cumulative value of the selected feature
- Associated events along this path is the storyboard

Interaction Mechanism 2

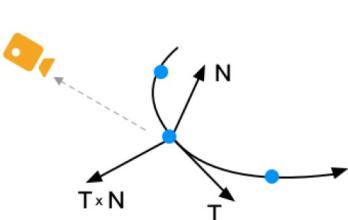
- Select the areas of interest
- Calculate path that passes through the specified area
- Associated events along this path is the storyboard



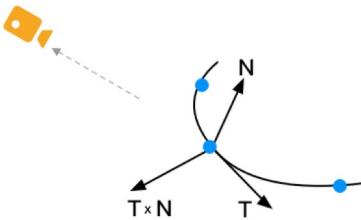
Shot Design Module

Why camera shot design

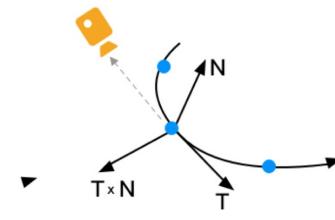
- Borrow shot designs in filmmaking
- Achieve a compelling and intuitive understanding
- Depict different features and correlation with the environment



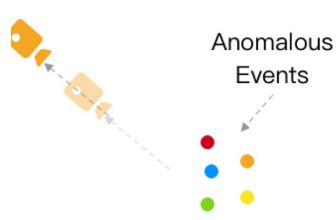
Normal shot
It tracks the objects from the side



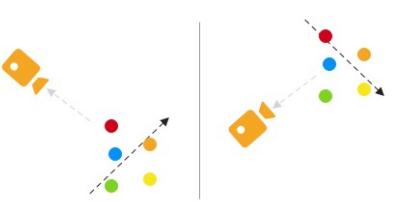
Panorama shot
It slowly moves the camera over a landscape with a wide-open space



Following shot
The camera follows the subject being filmed around



Dolly shot
It moves the camera toward or away from a subject while filming



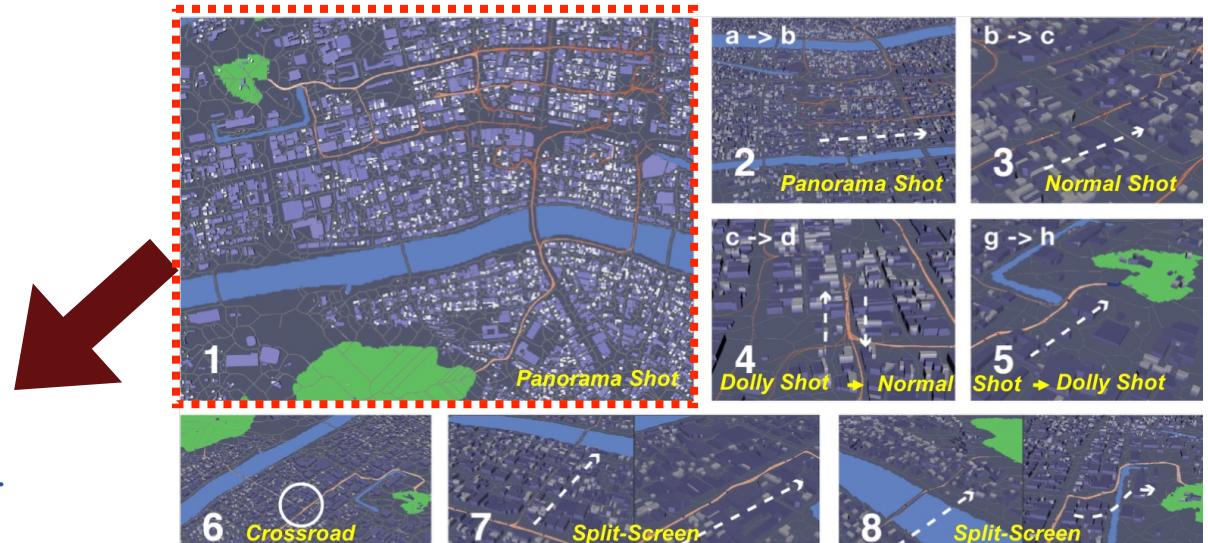
Split-screen
It divides the screen and shows several images simultaneously

Case: Exploring Evacuation by Using Shot Designs

Step 1. Specifying features of interest in evacuation

- Obtain a trajectory overview
- Specified an area of interest
- Compute the scene feature graph of the trajectories

Agents in the central area had to pass through some building-intensive areas to reach the shelter in the west



- Set turn-back as the preferred feature
- Optimal path automatically generated



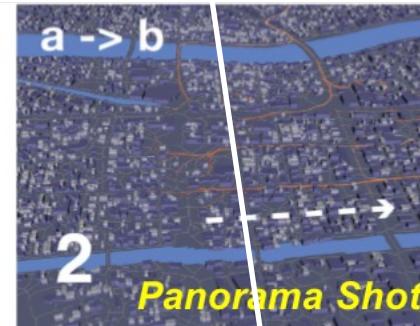
Case: Exploring Evacuation by Using Shot Designs

Step 2. Exploring correlation between features and environment

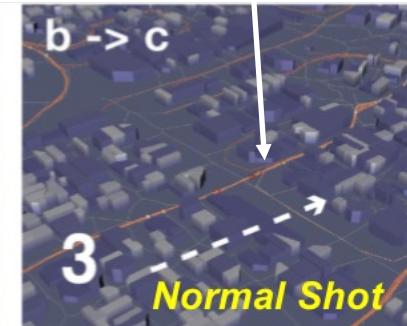
Tracking begins with a panorama shot



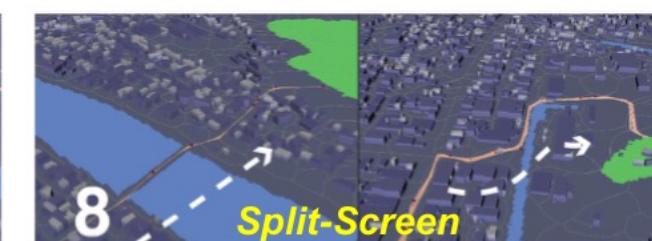
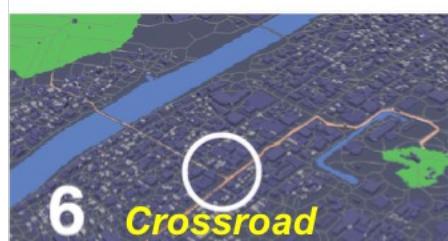
Detecting a turn-back event



Tracking the group

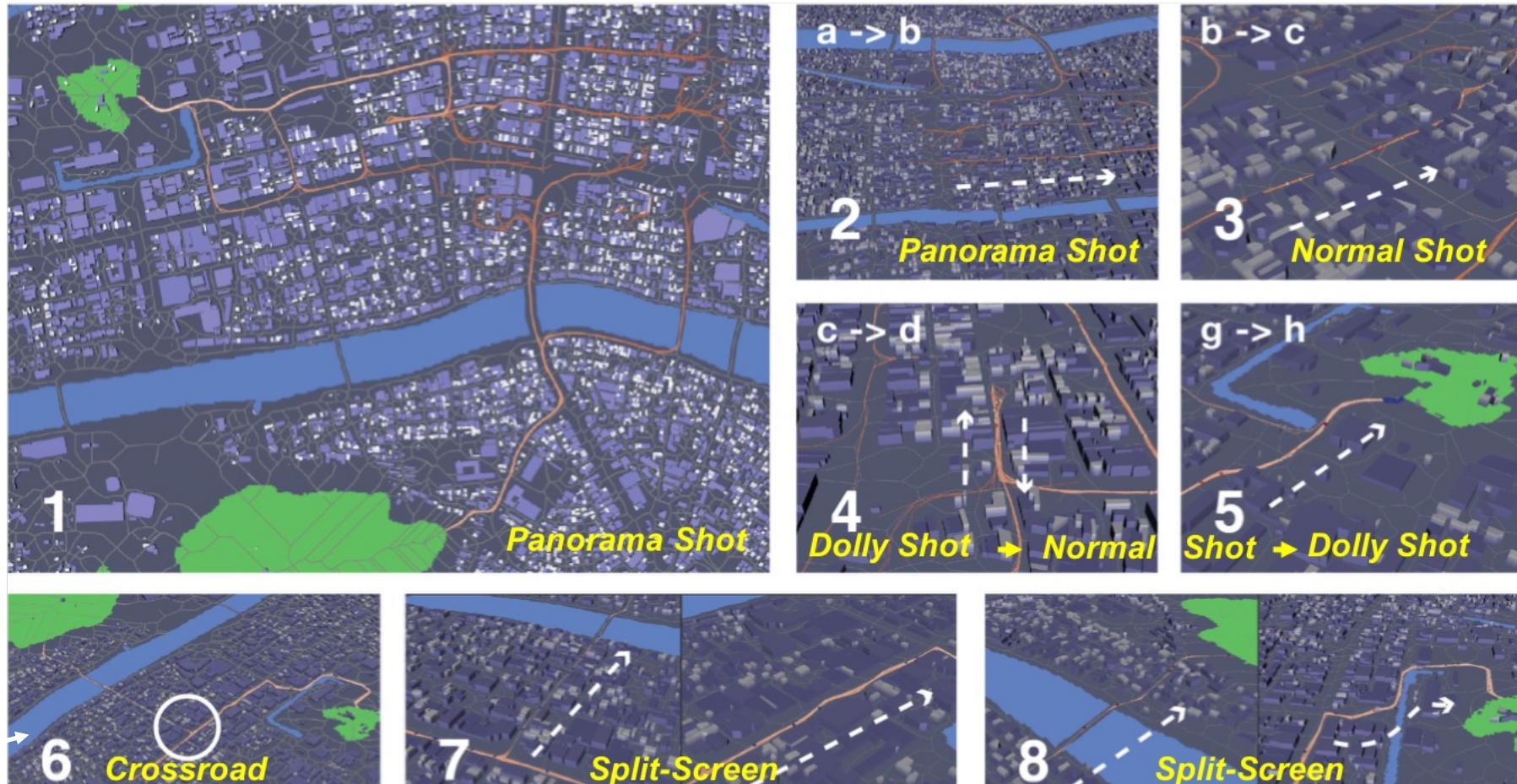


Continued filming the situation around the shelter



Case: Exploring Evacuation by Using Shot Designs

Step 3. Comparing different evacuation scenarios

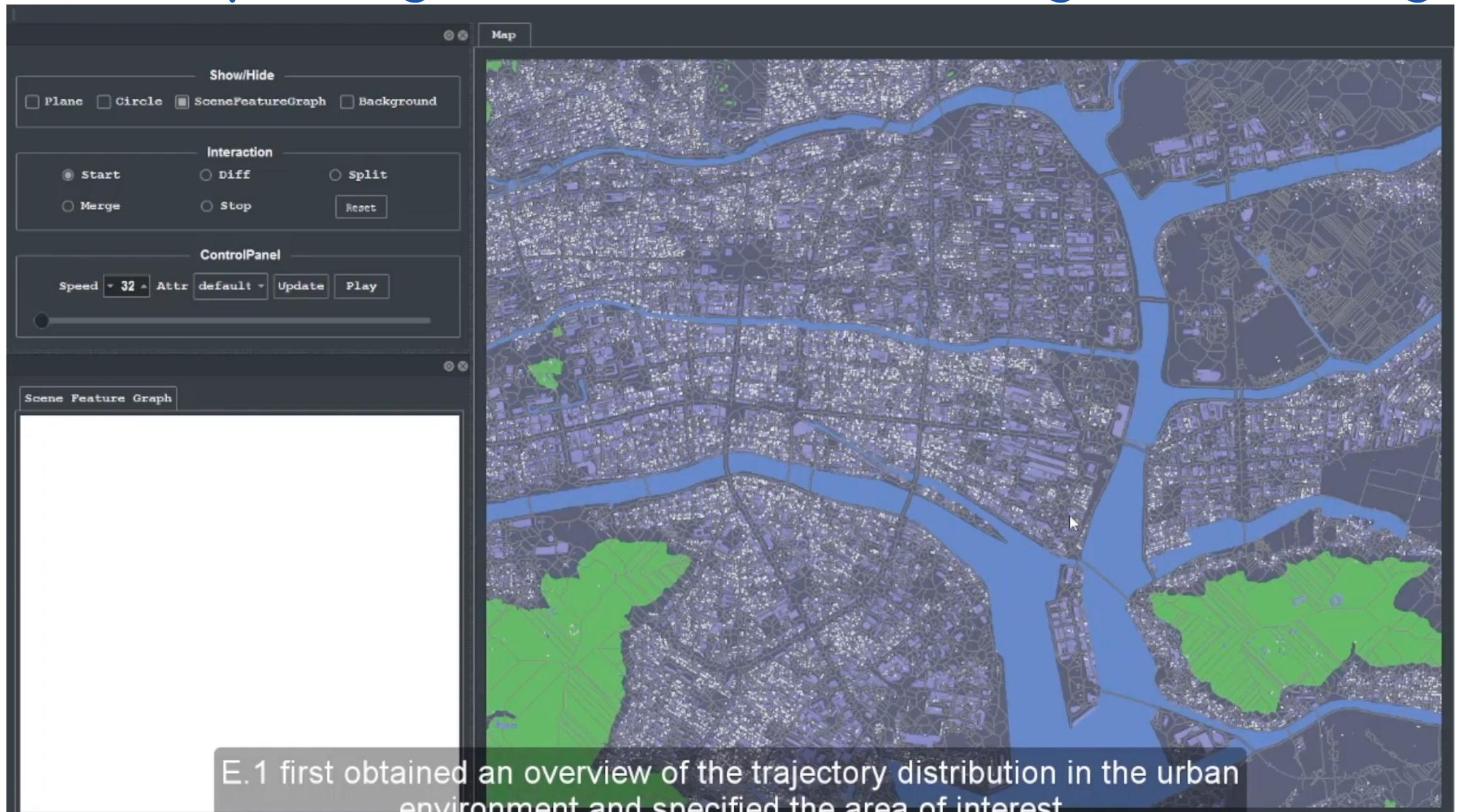


Upon arriving at the road intersection, some evacuees rushed to the shelter in the upper left corner while some fled to the shelter in the bottom right corner

Split the screen to track the two groups

The “split-screen” keeps tracking the two groups until they reached the shelters

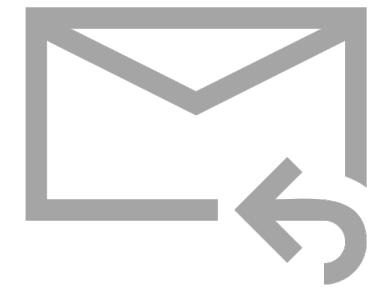
Case: Exploring Evacuation by Using Shot Designs





Quan Li

Questions?
Thank you 😊



liquan@shanghaitech.edu.cn