imMens

Real-time visual querying of big data

Limitations of traditional data visualization tools

- Inadequate to visualize big data

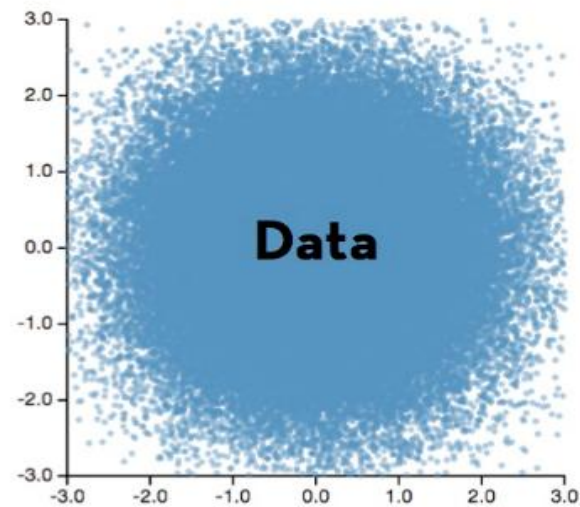- Perceptual Scalability

- Interactive Scalability

Two challenges

- Effective visual encoding
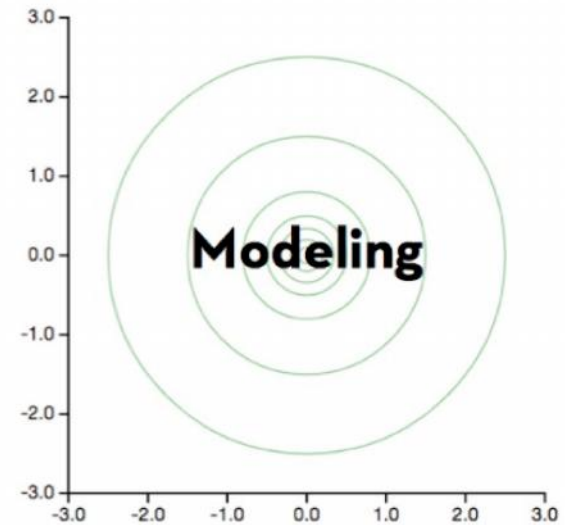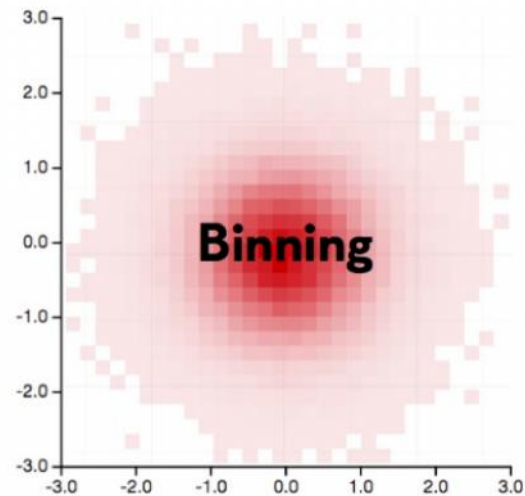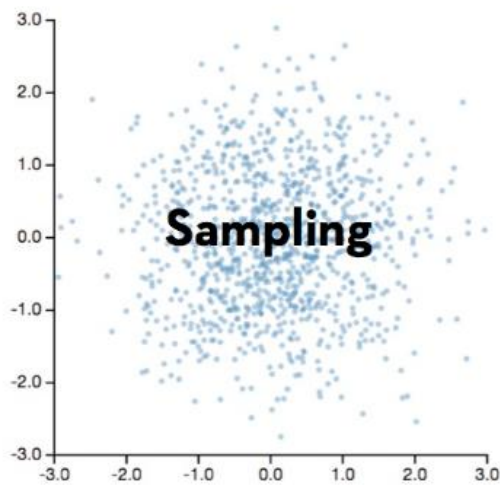
- Real-time interaction

Two challenges

- **Effective visual encoding**

- Real-time interaction

Perceptual scalability

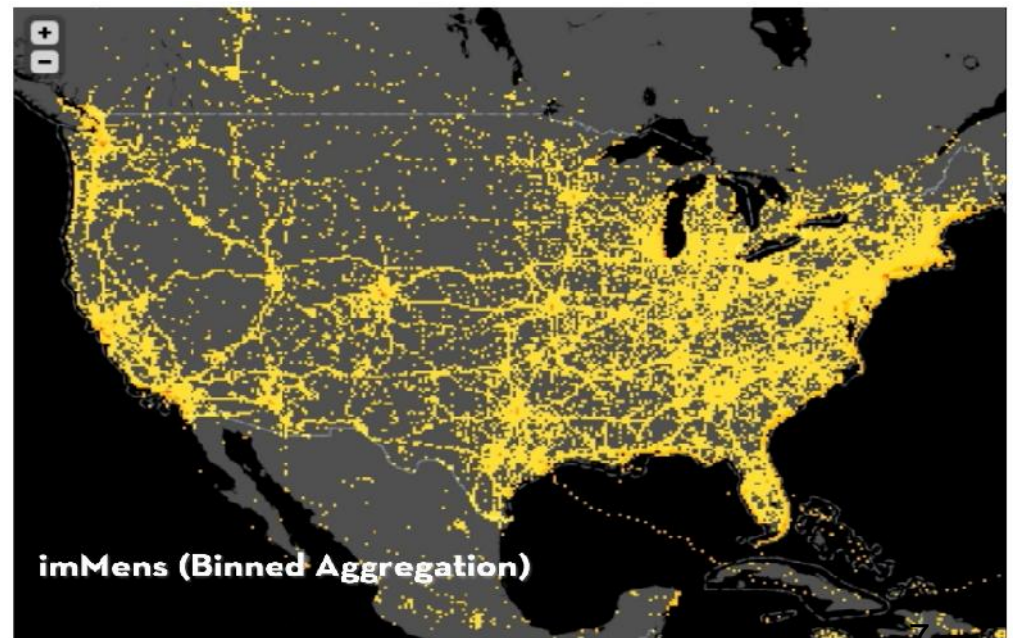How to create perceptually effective visualization?

Data reduction methods

Real data example

~over 4 million user check-ins on BrightKite, a location-based network service.

- Using binned aggregation

- Counting the number of points in each bin

- Using color to form a heatmap

imMens (Binned Aggregation)
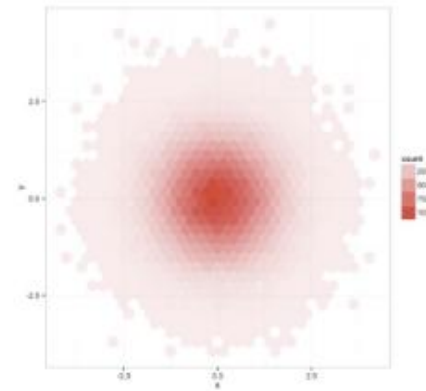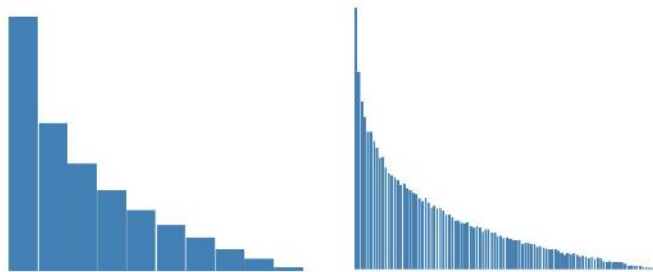
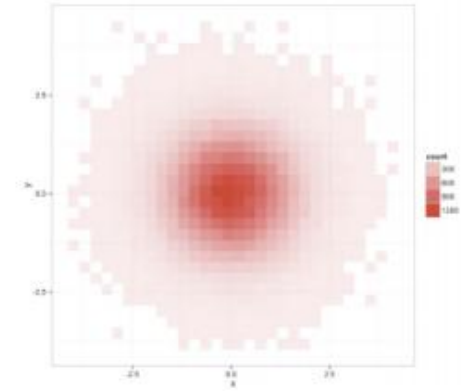# imMens General approach

Bin→ Aggregate → Plot

imMens General approach

Bin→ Aggregate → Plot

Number of Bins?

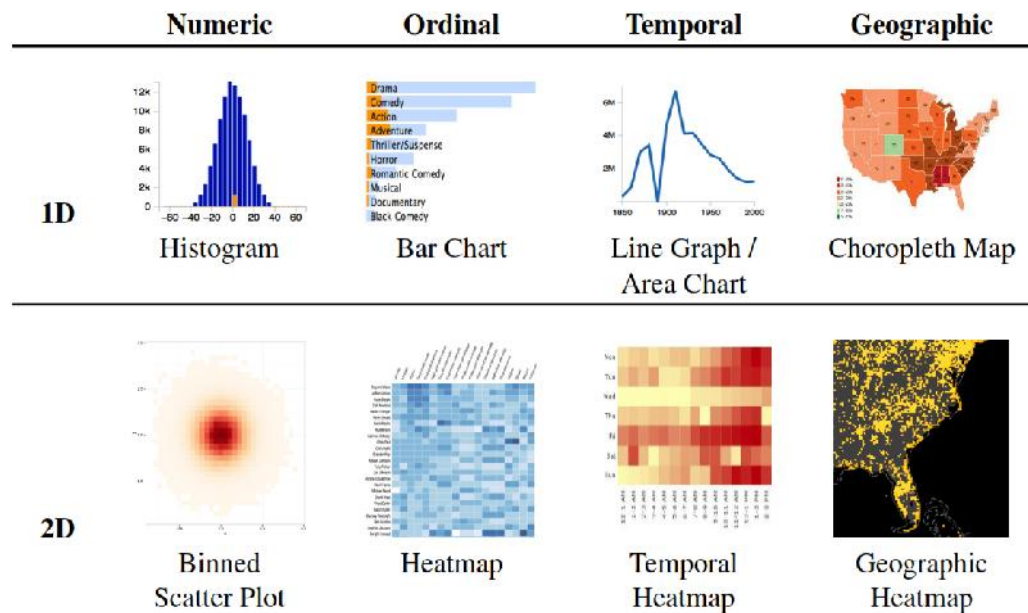Hexagonal Bins   Rectangular Bins

imMens General approach

Bin→ Aggregate → Plot

Aggregate : Count, Min, Max, Sum, Average

imMens General approach

Bin→ Aggregate → Plot

Plot:

	Numeric	Ordinal	Temporal	Geographic
1D	Histogram	Bar Chart	Line Graph / Area Chart	Choropleth Map
2D	Binned Scatter Plot	Heatmap	Temporal Heatmap	Geographic Heatmap

Two challenges

- Effective visual encoding

- **Real-time interaction**

Interactive Techniques

Select

    Details-on-demand

Navigate
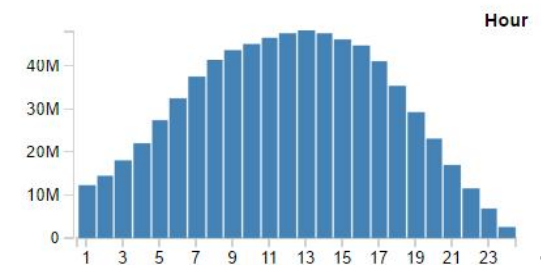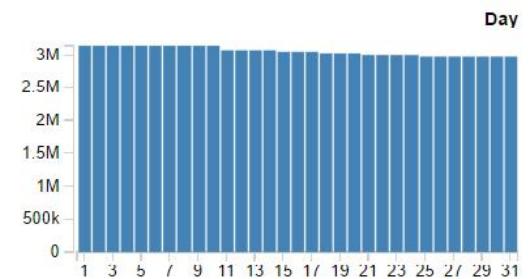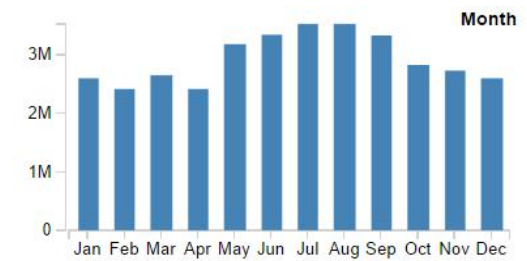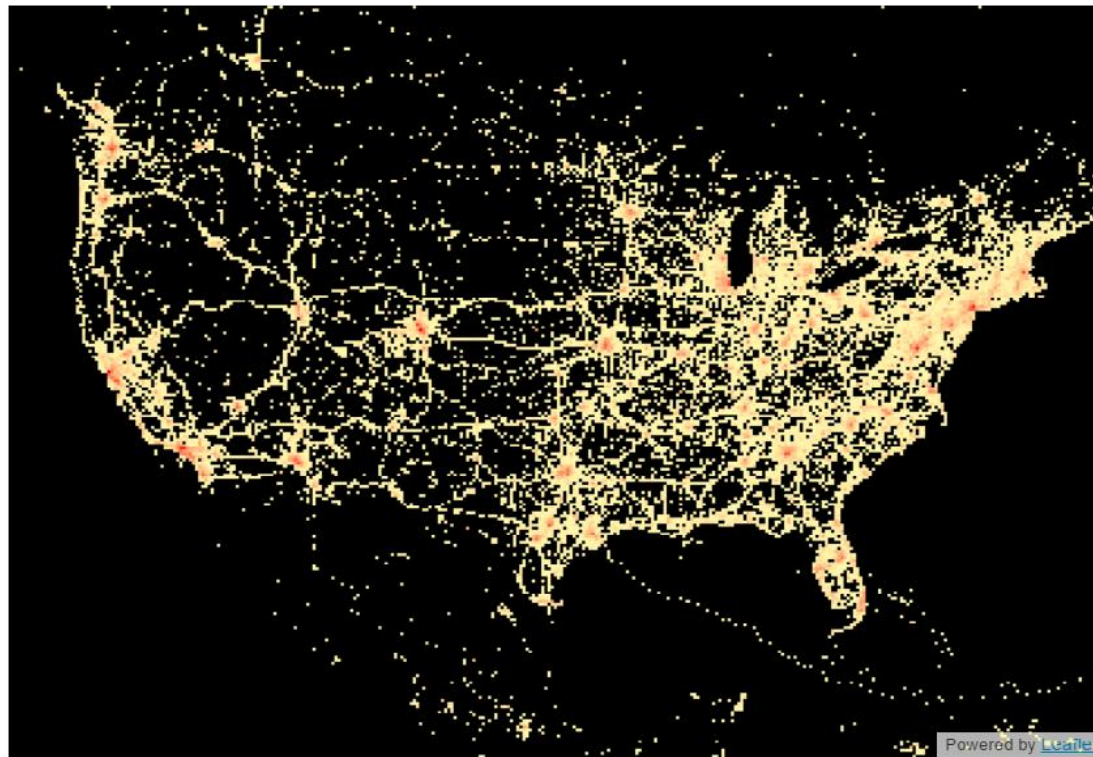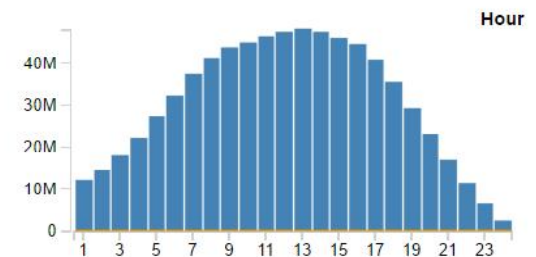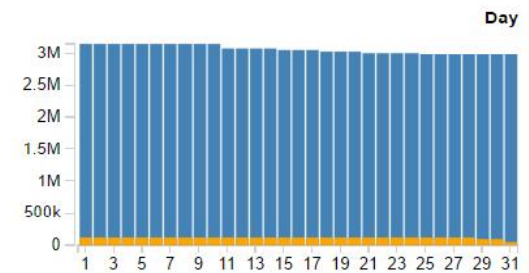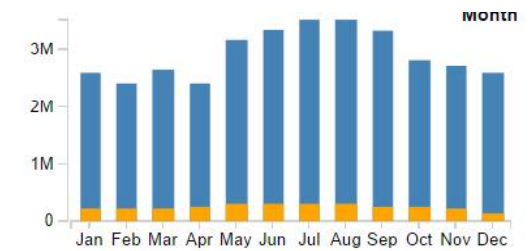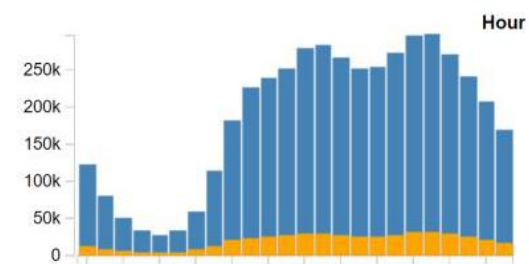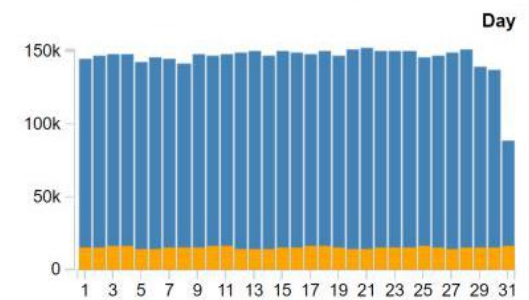
    Pan & Zoom
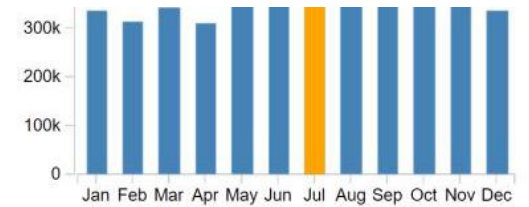
Query

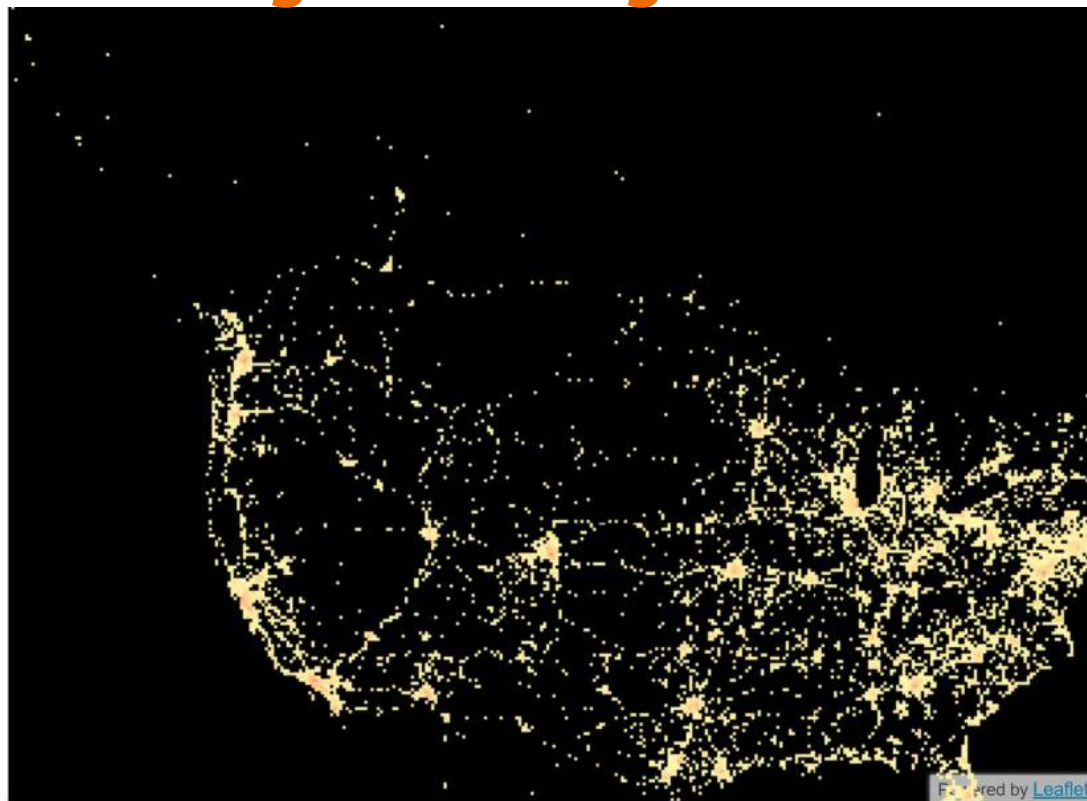   Brush & Link (difficult to support)

Brightekite

Select & Navigate

Brushing & Linking

16

Data cube query to support interaction

5-D Data Cube
Month, Day, Hour, X, Y
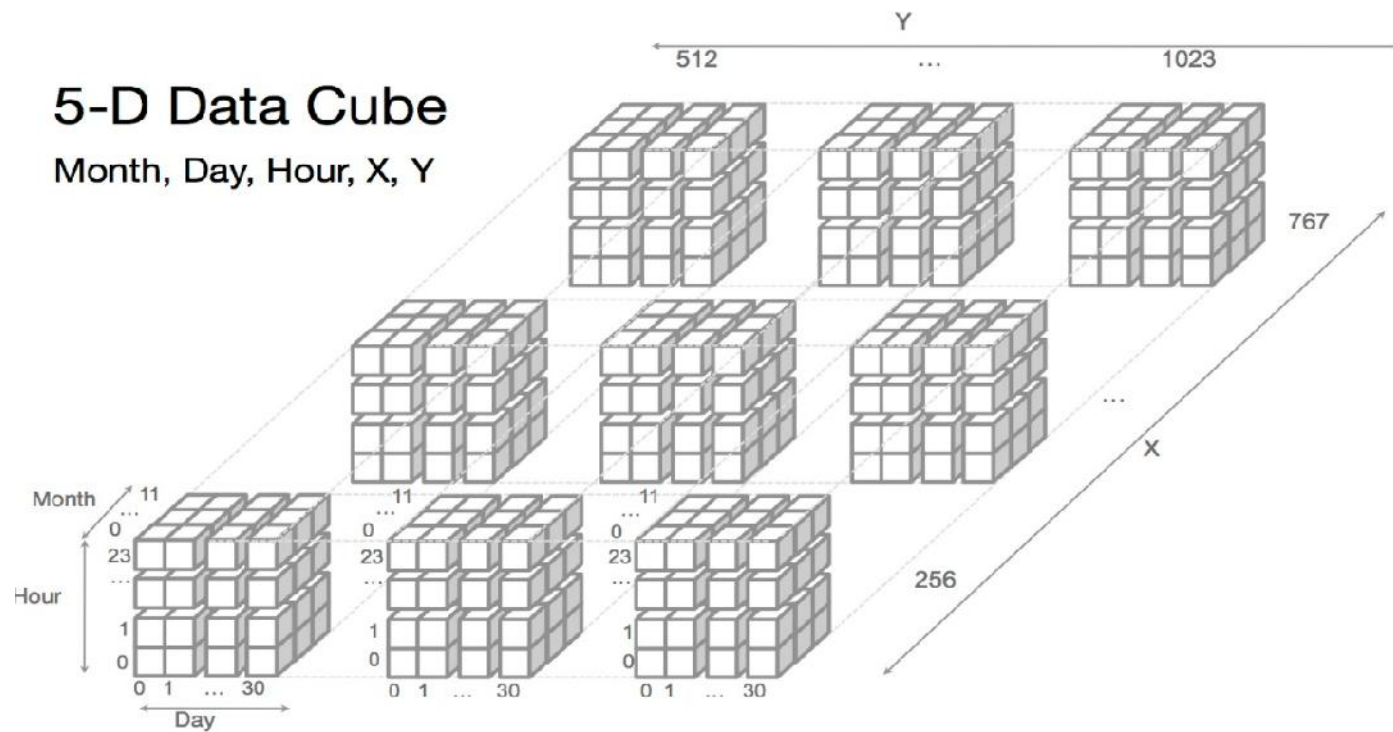
12 x 31 x 24 x 512 x 512 = ~**2.3 billion cells**

Data cube query to support interaction

Brushing January
Month, Day, Hour, X, Y
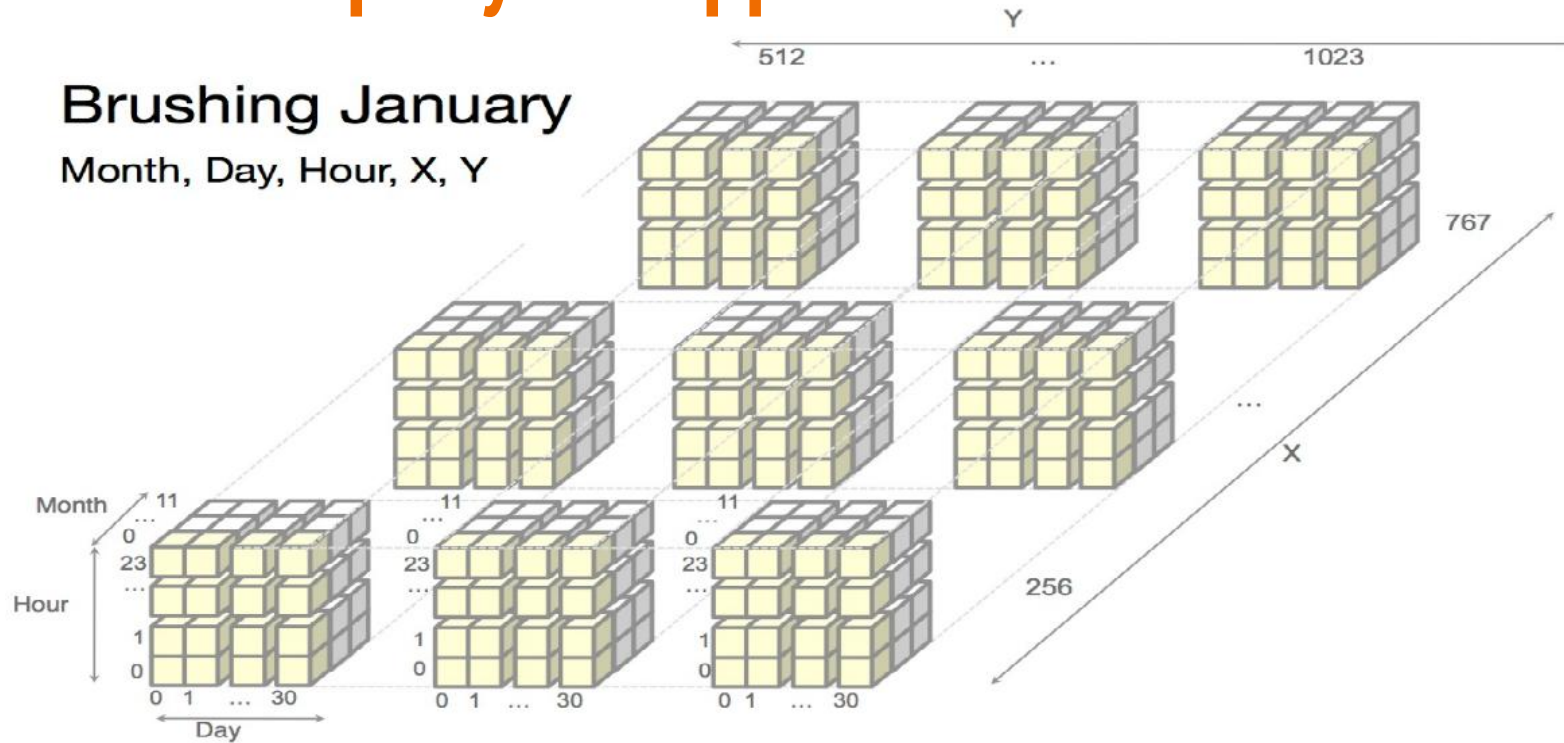
31 x 24 x 512 x 512 = ~**195 million cells**

Decomposing into sub-cubes
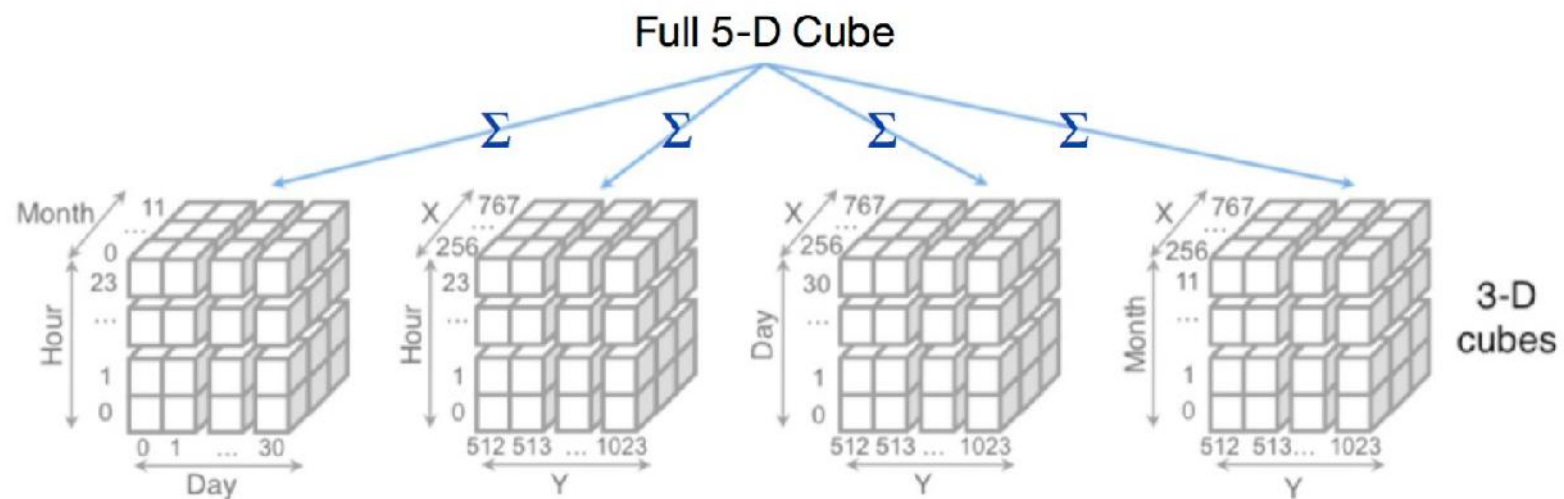
imMens: From data cubes to multivariate data tiles

1.Send raw data, not pixel

- Able to query and analyze

2.Embed multi-dim data

- Support brushing and linking
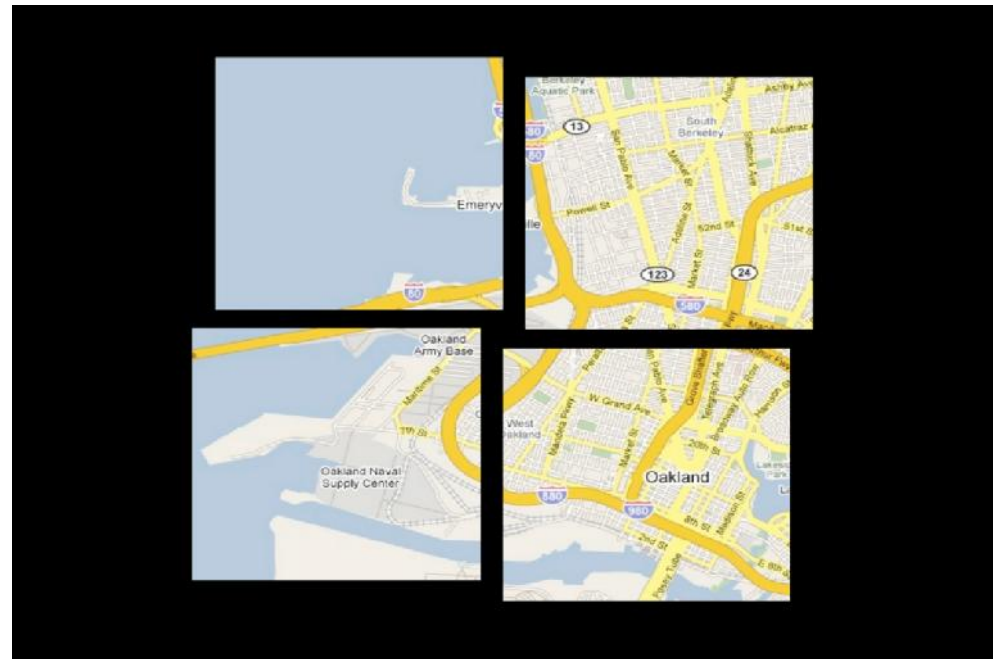
Google Map

The sense of Data reduction

The number of dimensions: **5 (full 5-D cube)**
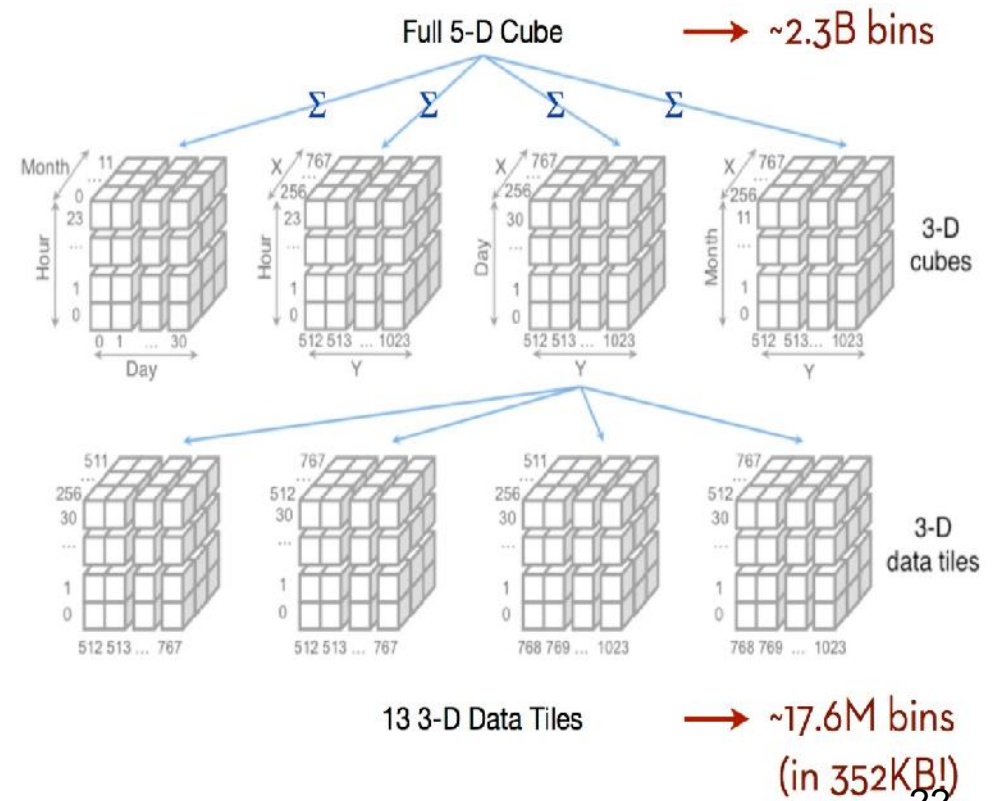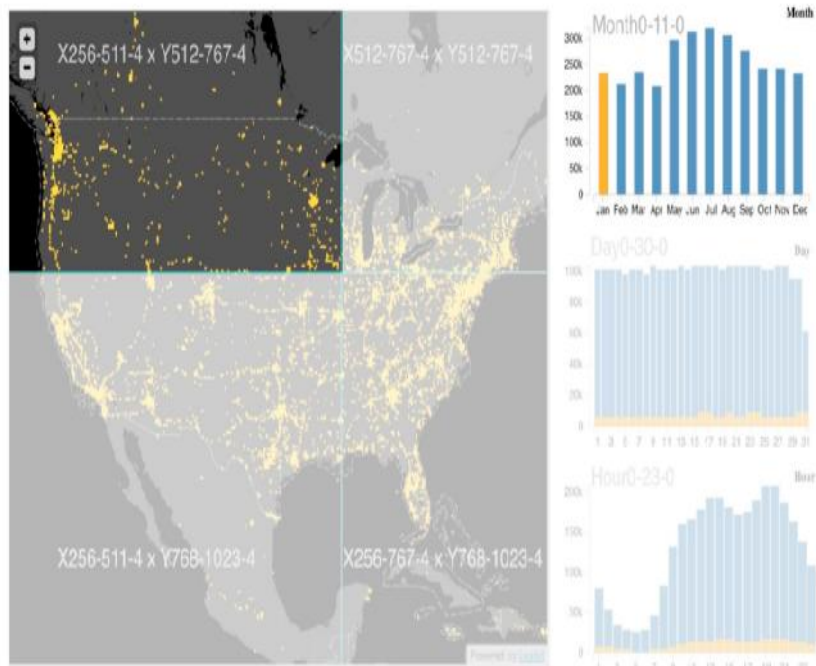
Uniform bin count = **b**

The number of data records is = ,

Data tile decomposition ➜
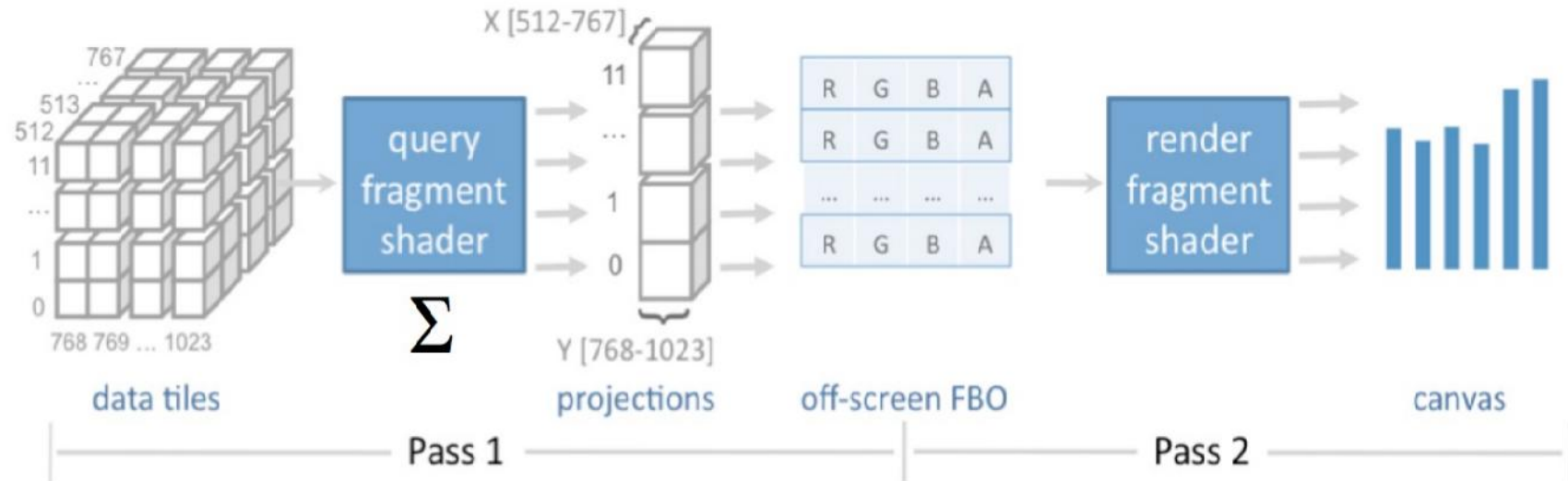
The number of records is = ,

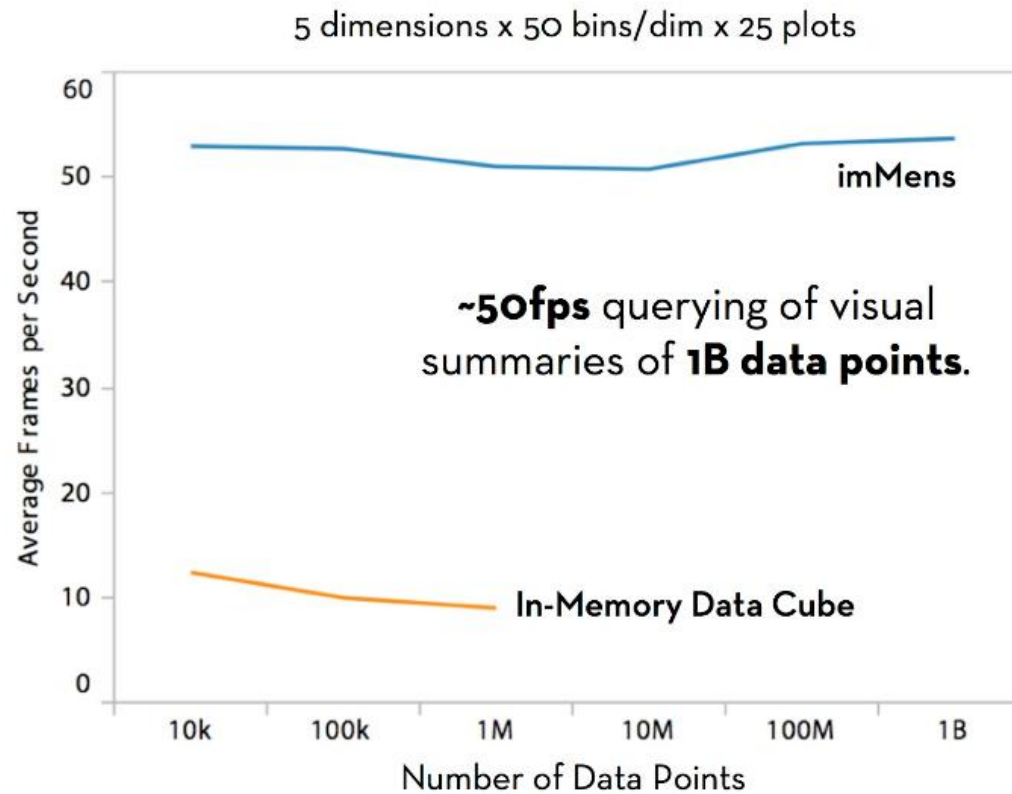*b=50* ➜ # of records is reduced from **312.5M** to **0.5 M** records

Brightkite!

Parallel query processing

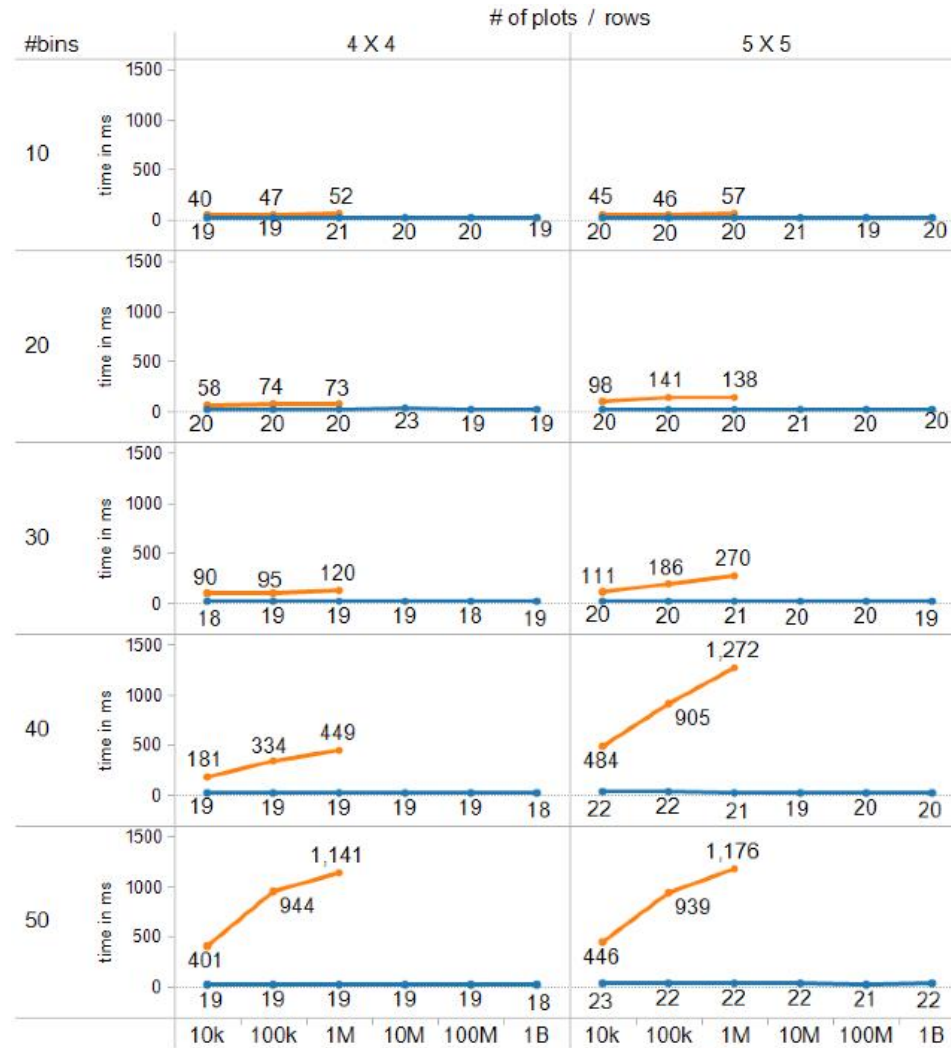# Results

5 dimensions x 50 bins/dim x 25 plots

~**50fps** querying of visual summaries of **1B data points**.

Results

25

Results

Data Set	Brightkite	Flight Delays	SPLOM
Size	4M	118M	1B
Bins	Month (12)	Carrier (28)	Dim. A (50)
	Day (31)	Year (20)	Dim. B (50)
	Hour (24)	Day of Week (7)	Dim. C (50)
	X (256)	Dep. Delay (174)	Dim. D (50)
	Y (256)	Arr. Delay (174)	Dim. E (50)
Data Tiles	13	4	10
Time	*17.76 ms*	*16.56 ms*	*20.47 ms*

Summary

- We contribute methods for real-time visual querying of big data.

- We integrate multivariate data tiles and parallel processing.

- We use WebGL to access GPU processing in a web browser.

- First system to enable real-time interaction with data sets this large.

Thank You!