Data Science
SONYC

NYC, February 8, 2018

Fabio Miranda

PhD Candidate

Together with Harish Doraiswamy, Yitzchak Lockerman, Marcos Lage, Charlie Mydlarz, Justin Salamon, Yurii Piadyk, Fernando Chirigati, Juliana Freire, and Claudio T. Silva

NYU

VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

SONYC Data Science

- Analysis of SONYC data – 34 years worth of data
- Analysis of SONYC together with multiple data sets
  - E.g.: How construction permits impact SPL captured by SONYC
- Data collected from traditional and *unsuspecting* sensors
  - SONYC, census, crime, building permits, public transportation, tweets

Opportunity: leverage this data to make new insights about how people are using cities, frame new policies and make cities more efficient

Challenges of Data Science

- SONYC: 34 years worth of data
  - How to handle and query large data?
  - How to visualize this data?
  - How to gain new insights from the data?

Objectives for SONYC

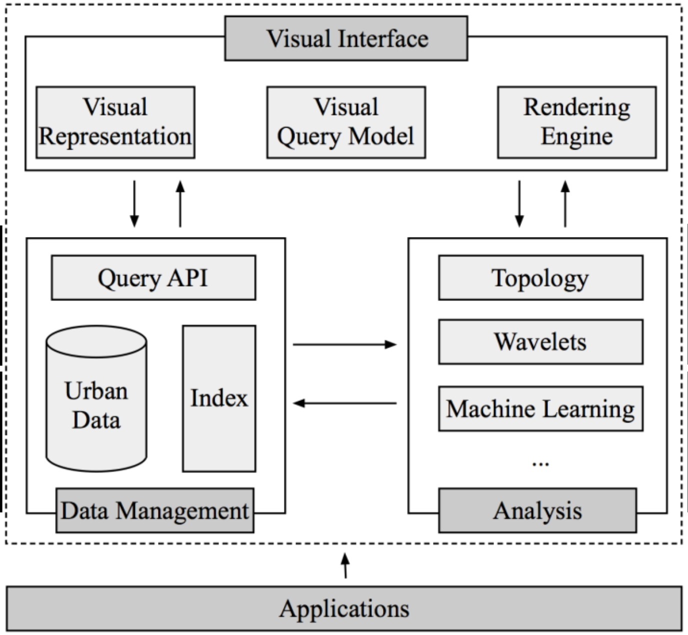
- Interactive querying of noise data
  - Techniques to support interactive, low latency queries of SPL data
  - Drive exploratory visualization
- Visual interface
  - Build a visual interface for noise data exploration
  - Explore noise in the context of the city and related data
- Analysis of city-wide noise
  - Data analytics to gain insights into possible patterns of noise over space and time
  - Use the generated data (SPL) together with open data
  - Generate a city-wide time-varying noise map

Vision

Noise Profiler

STIG [ICDE '16]
Raster Join [PVLDB '17]

Time Lattice

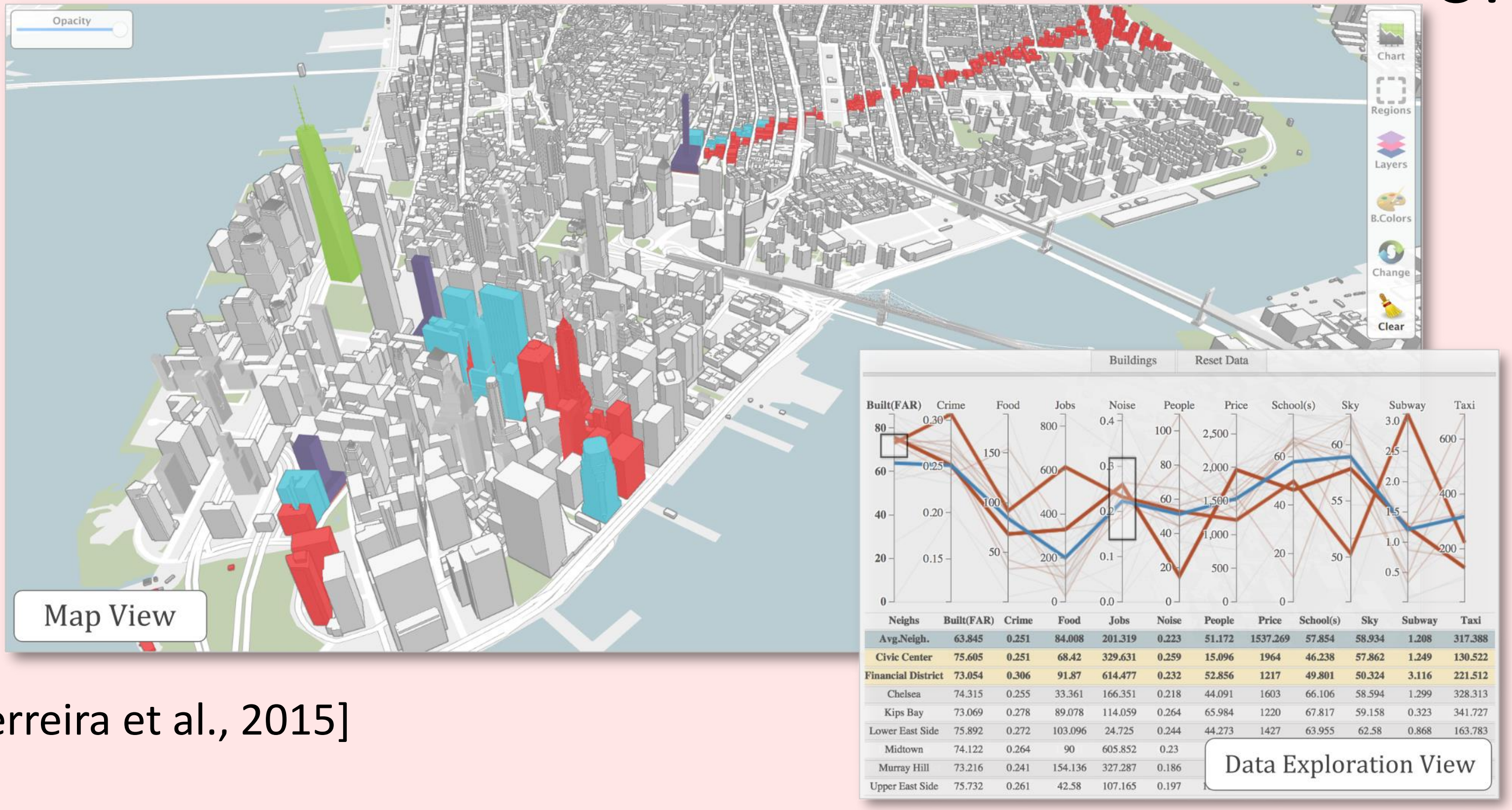


Urban Pulse [VIS '16]
Data Polygamy [SIGMOD '16]
Shadow Accrual Maps [TVCG '18]

AHV Analysis
Sound Simulation

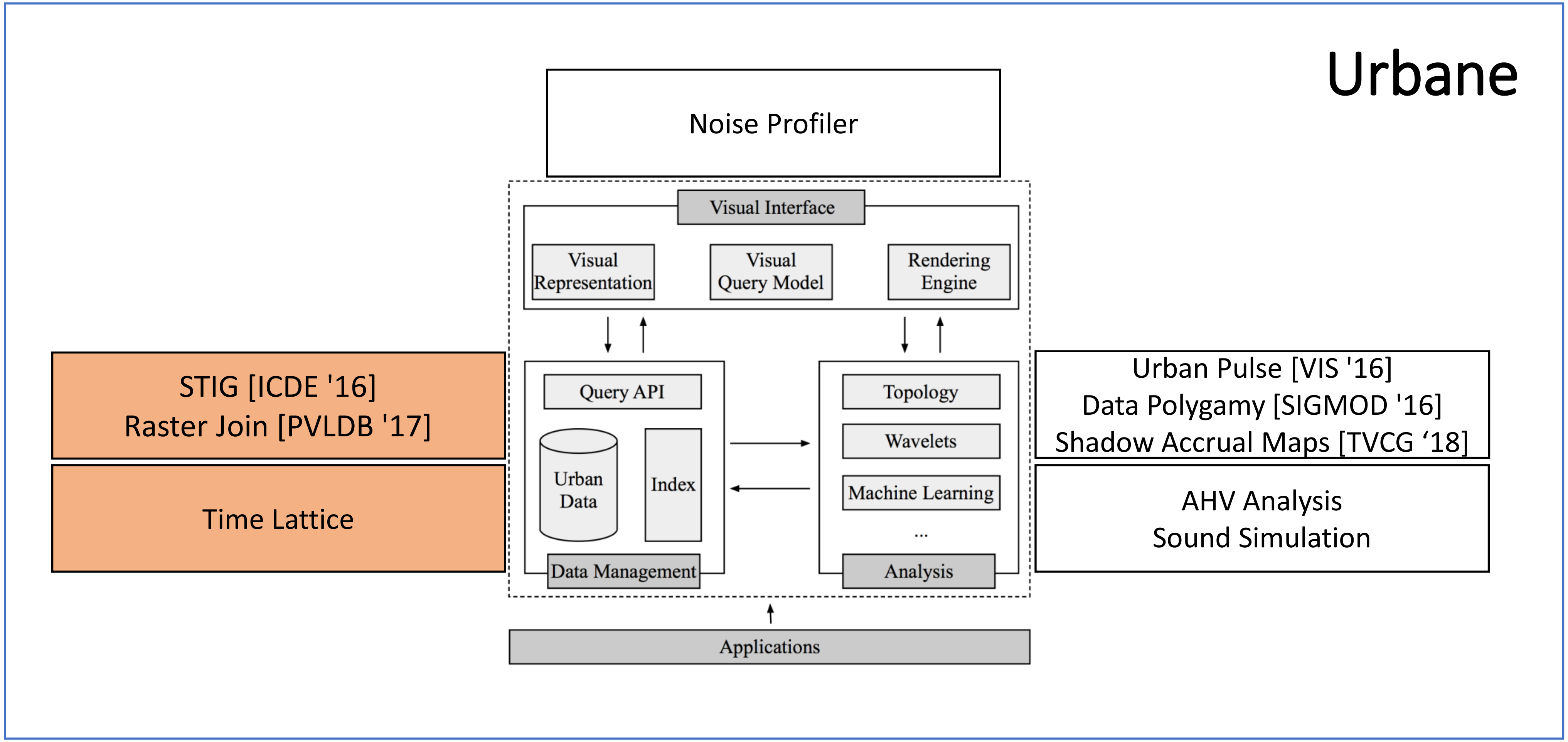
Vision

Urbane



[Ferreira et al., 2015]

Vision


Urbane
Noise Profiler
Visual Interface
Visual Representation
Visual Query Model
Rendering Engine
STIG [ICDE '16]
Raster Join [PVLDB '17]
Time Lattice
Query API
Urban Data
Index
Data Management
Topology
Wavelets
Machine Learning
...
Analysis
Urban Pulse [VIS '16]
Data Polygamy [SIGMOD '16]
Shadow Accrual Maps [TVCG '18]
AHV Analysis
Sound Simulation
Applications

Handling Large Temporal Noise Data

- Objective
  - Support queries having constraints at multiple time resolutions
    - Average SPL each hour of the day
    - Average SPL day of the week
    - Average SPL each day of the week, between 8am – 6pm
  - Support range queries at multiple resolutions
    - Average SPL between March 1st and March 15th, at hour resolution
  - Support updates from new data

Handling Large Temporal Noise Data

	Size		Q1		Q2		Q3		Q4	
	(MB)	Overhead	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup
Nanocube	41799	10349 %	116		4.6		2491.8		40083	
Pandas	1600	300 %	1670		9355		10399		11070	
InfluxDB	412	3%	10574		42913		35259		29058	
TimescaleDB	7867	1866%	20385		60206		130594		101036	
KairosDB	1301	225%	229110		629886		240168		75267	

NYU

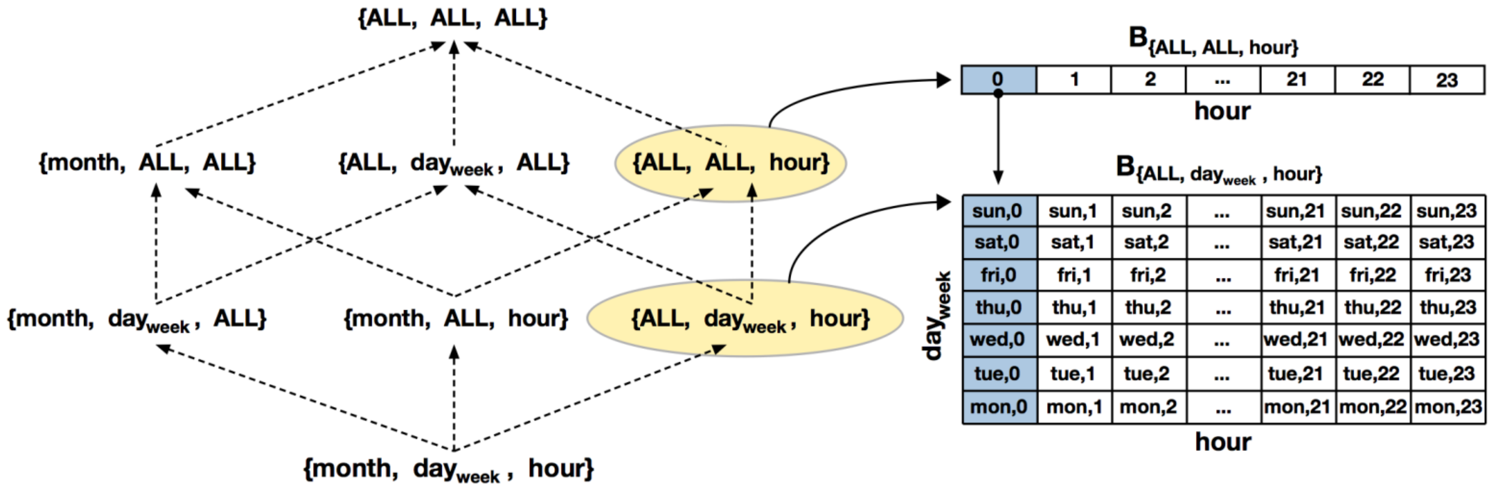
VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Handling Large Temporal Noise Data

- Objective
  - Support queries having constraints at multiple time resolutions
    - Average SPL each hour of the day
    - Average SPL day of the week
    - Average SPL each day of the week, between 8am – 6pm
  - Support range queries at multiple resolutions
    - Average SPL between March 1st and March 15th, at hour resolution
  - Support updates from new data
  - **Small memory** overhead
  - Allow low latency queries over large time series **(< 1 second)**

Handling Large Temporal Noise Data

- Time Lattice
  - Data structure that supports multiple resolution queries at interactive rates
  - Makes use of the implicit hierarchy present in temporal resolutions to materialize a sub-lattice of a data cube



Handling Large Temporal Noise Data

	Size		Q1		Q2		Q3		Q4	
	(MB)	Overhead	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup
Nanocube	41799	10349 %	116		4.6		2491.8		40083	
Pandas	1600	300 %	1670		9355		10399		11070	
InfluxDB	412	3%	10574		42913		35259		29058	
TimescaleDB	7867	1866%	20385		60206		130594		101036	
KairosDB	1301	225%	229110		629886		240168		75267	
Time Lattice	407	1.75%	40	-	15	-	12	-	92	-

NYU

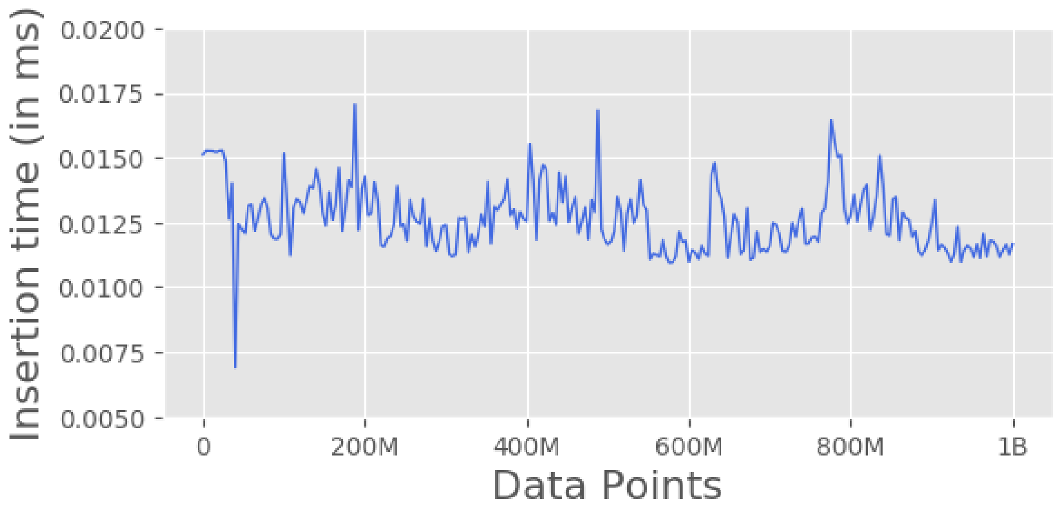
VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Handling Large Temporal Noise Data

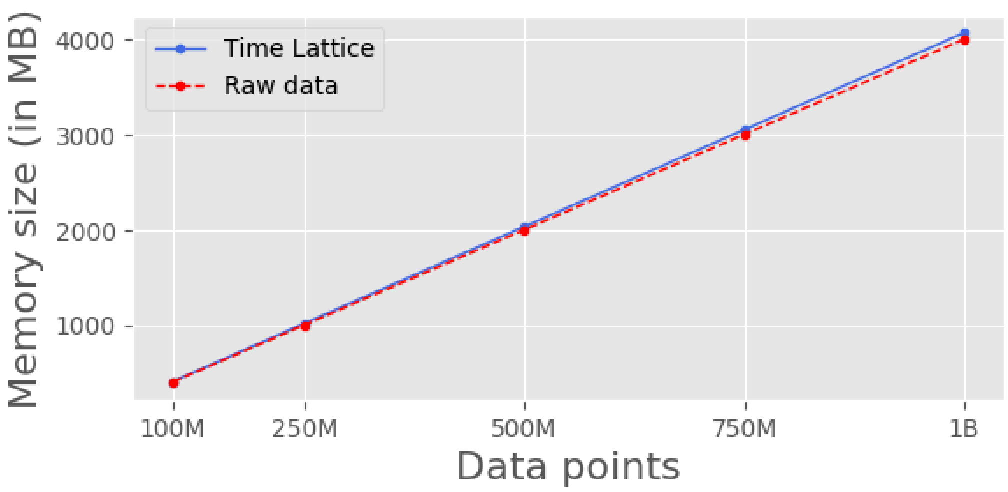
	Size		Q1		Q2		Q3		Q4	
	(MB)	Overhead	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup	Time(ms)	Speedup
Nanocube	41799	10349 %	116	2.9x	4.6	0.3x	2491.8	194x	40083	433x
Pandas	1600	300 %	1670		9355		10399		11070	
InfluxDB	412	3%	10574	261X	42913	2860x	35259	2754x	29058	314x
TimescaleDB	7867	1866%	20385		60206		130594		101036	
KairosDB	1301	225%	229110		629886		240168		75267	
Time Lattice	407	1.75%	40	-	15	-	12	-	92	-

VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Handling Large Temporal Noise Data



Constant insertion time: ideal for streaming



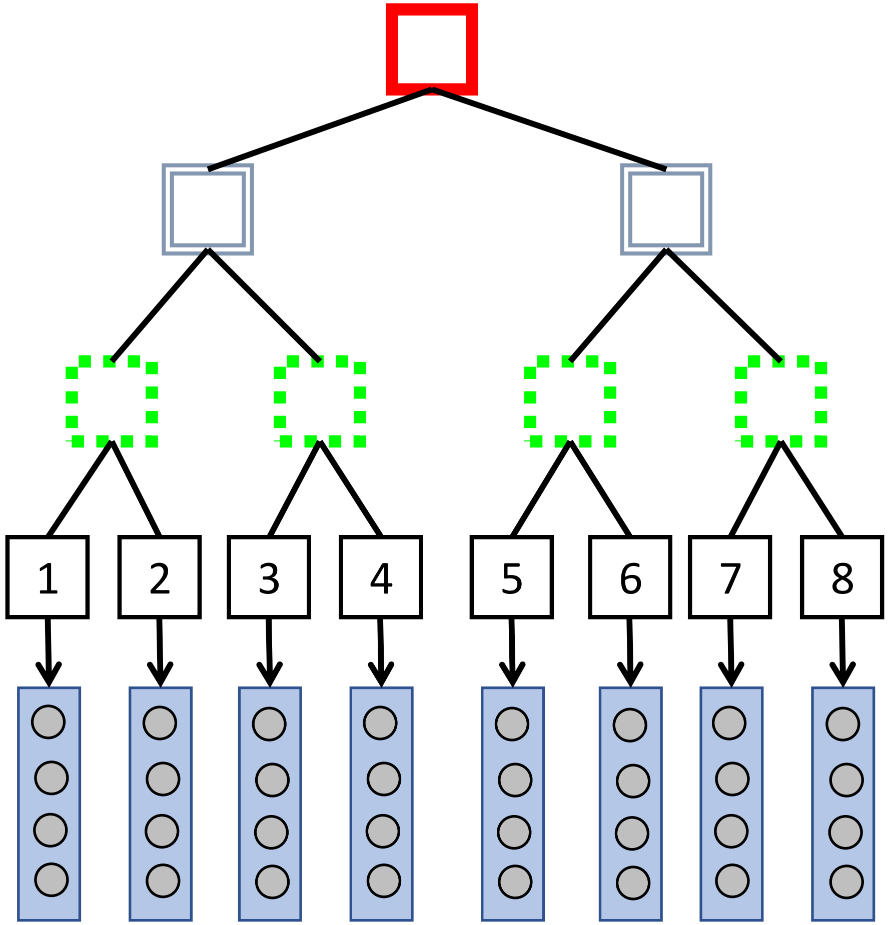
Linear memory overhead

Handling Large Spatio-Temporal Data

- Developing a set of GPU-based techniques
- STIG [Doraiswamy et al. 2015]

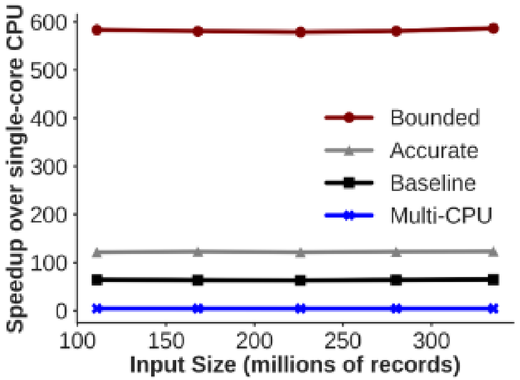
Query	MongoDB	PostgreSQL		ComDB	
	Time	Time	Speedup	Time	Speedup
1	0.075	503.9	6718x	20.6	274x
2	0.080	501.9	6273x	23.3	291x
3	0.067	437.8	6534x	21.6	322x
4	0.070	437.1	6244x	32.6	465x

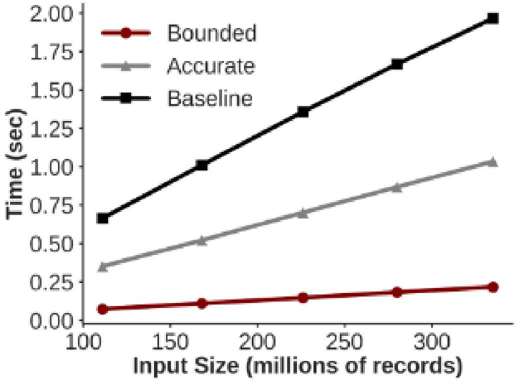
Time in Seconds

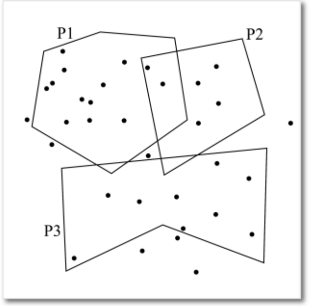


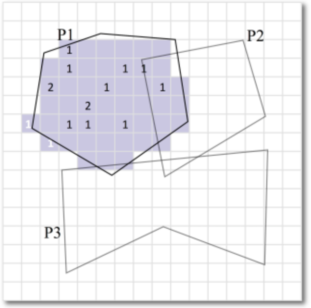
Handling Large Spatio-Temporal Data

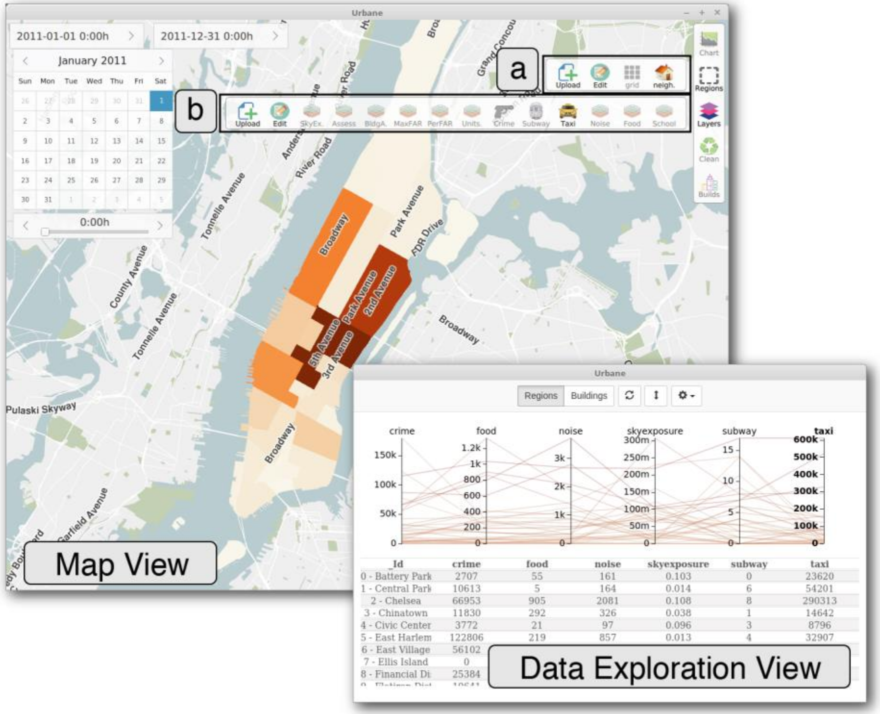
- Raster join [Tzirita Zacharatou, Doraiswamy et al., 2017]









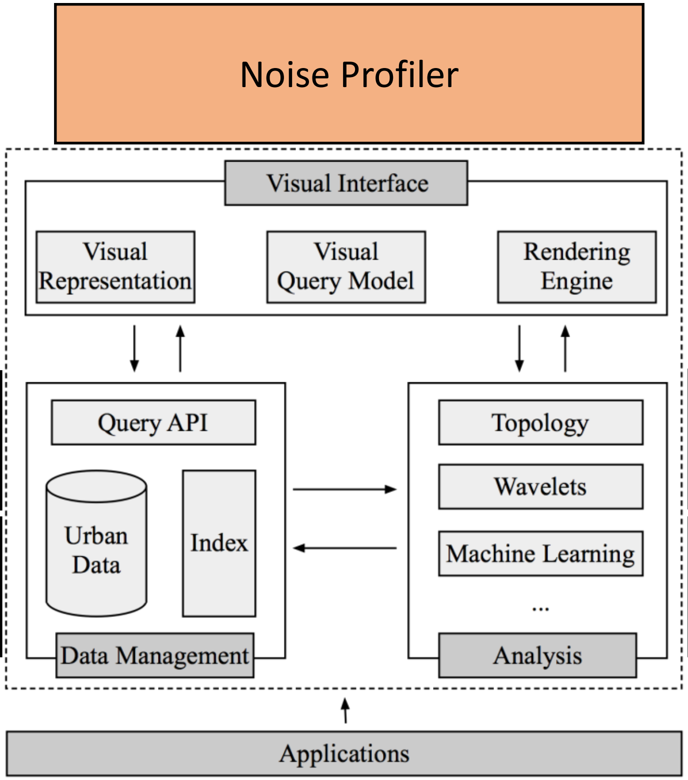


Vision

Urbane

STIG [ICDE '16]
Raster Join [PVLDB '17]

Time Lattice



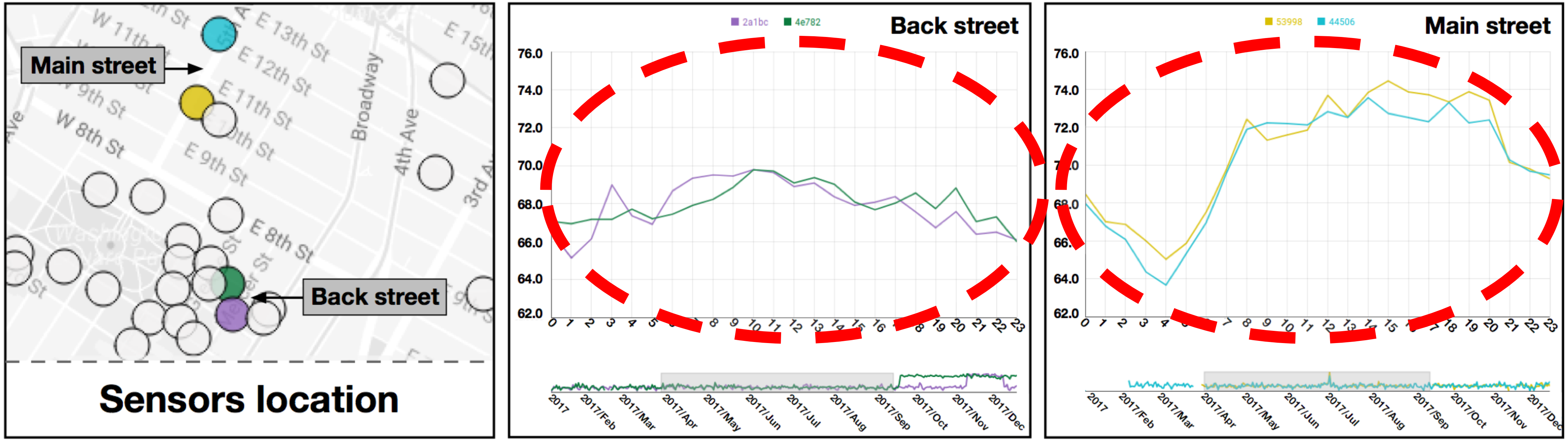
Urban Pulse [VIS '16]
Data Polygamy [SIGMOD '16]
Shadow Accrual Maps [TVCG '18]

AHV Analysis
Sound Simulation

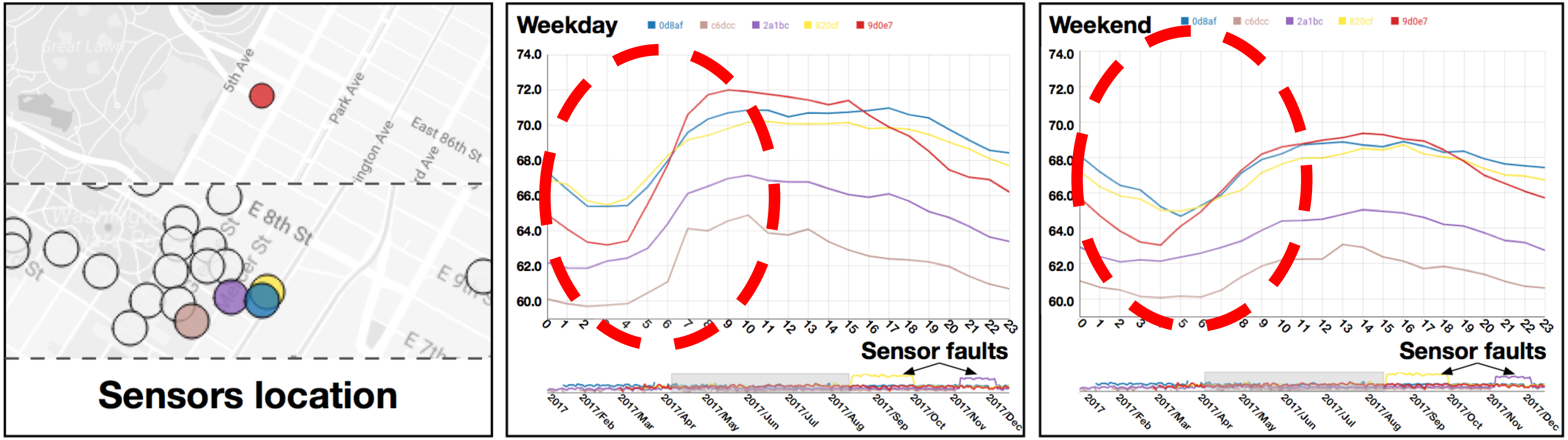
Time Lattice Interface: Noise Profiler

- Noise Profiler
  - Enable domain experts to specify, execute and visualize queries over the SPL data from across the city.
  - Compare data from one or more sensors
  - Support multiple metrics as the aggregate in the queries (e.g. equivalent continuous A-weighted sound pressure level)

Time Lattice Interface: Noise Profiler



Time Lattice Interface: Noise Profiler



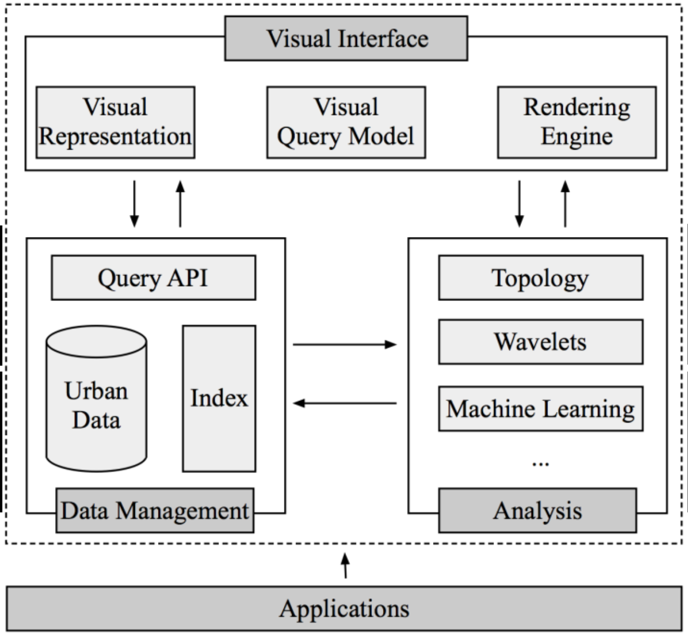
NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Vision

Urbane

Noise Profiler



STIG [ICDE '16]
Raster Join [PVLDB '17]

Time Lattice

Urban Pulse [VIS '16]
Data Polygamy [SIGMOD '16]
Shadow Accrual Maps[TVCG ‘18]

AHV Analysis
Sound Simulation

Analysis of after hour variances

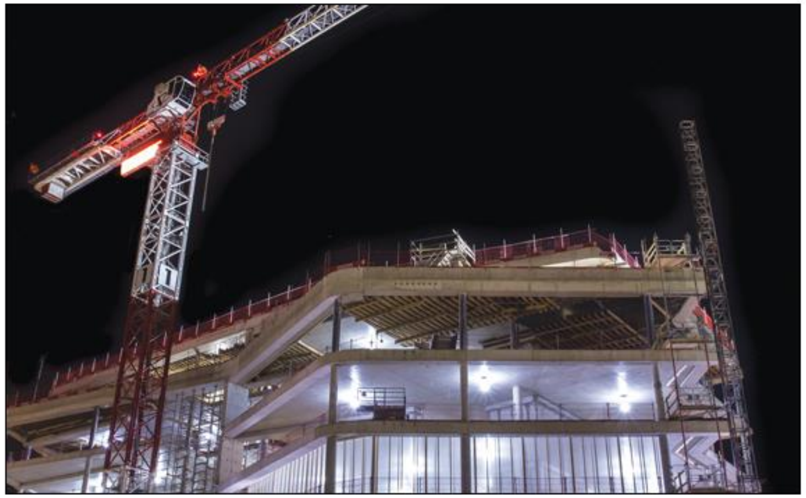
THE REAL DEAL
NEW YORK REAL ESTATE NEWS
New York Miami Los Angeles
Search

April 2016 Issue

The after-hours construction boom

Why buildings are rising on nights and weekends

By Kathryn Brenzel | April 01, 2016 12:00PM



(Photo: Shutterstock)

NEW YORK POST

Buildings Dept. approves night construction, angering residents

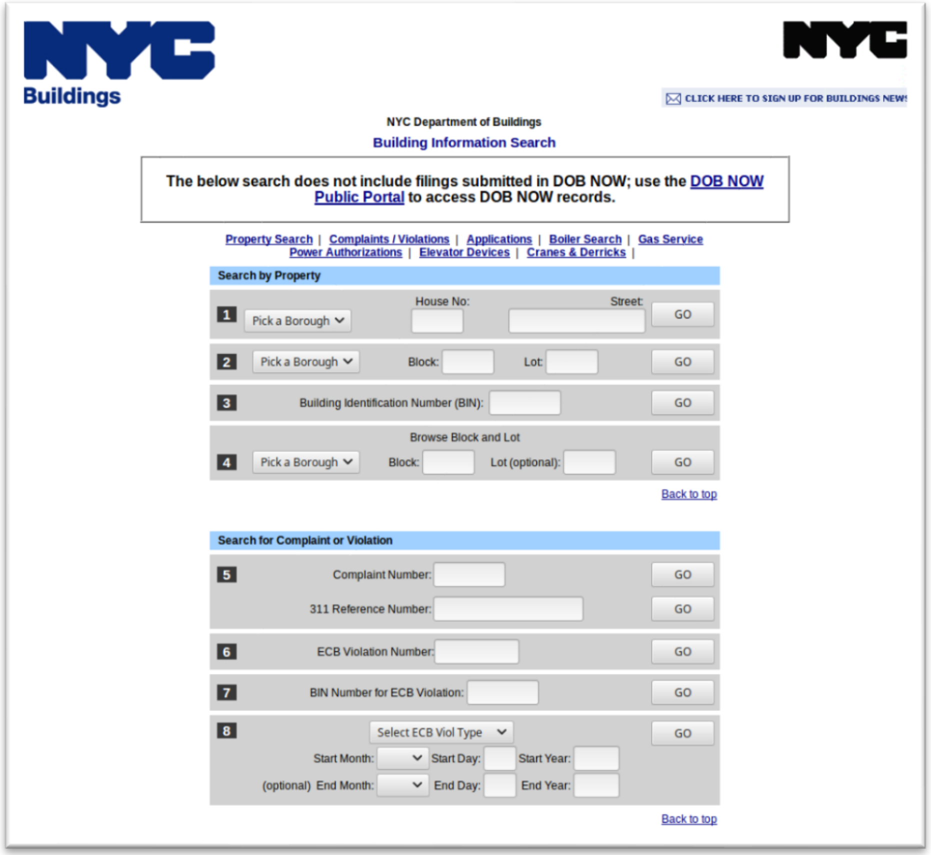
By Isabel Vincent and Melissa Klein January 31, 2016 | 1:51am

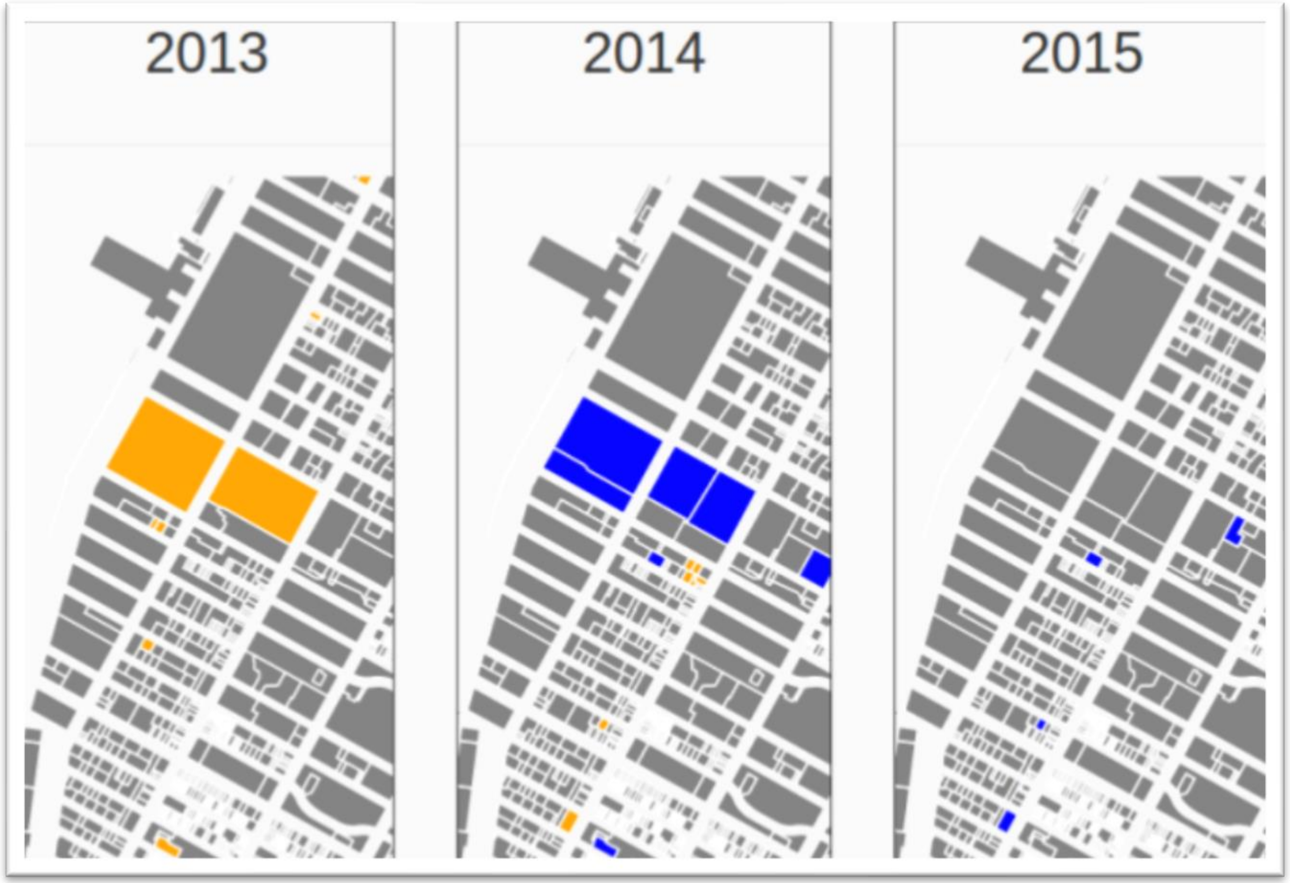
Shutterstock

NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

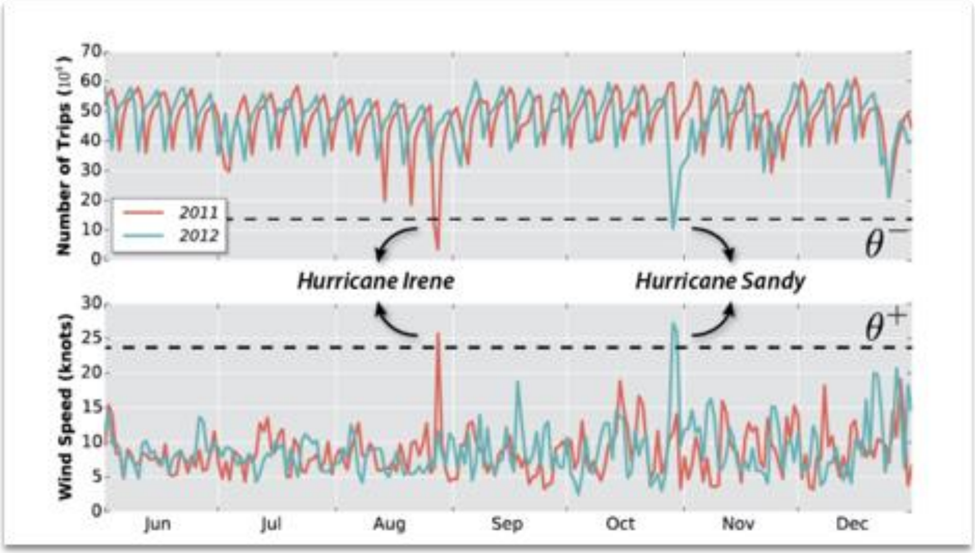
Analysis of after hour variances





Find spatio-temporal relationships

- Data Polygamy [Chirigati et al., 2016]
  - 100's of spatio-temporal data sets
  - Relationships occur only over certain points in space and time
  - Millions of possibilities
  - How to efficiently identify interesting relationships?



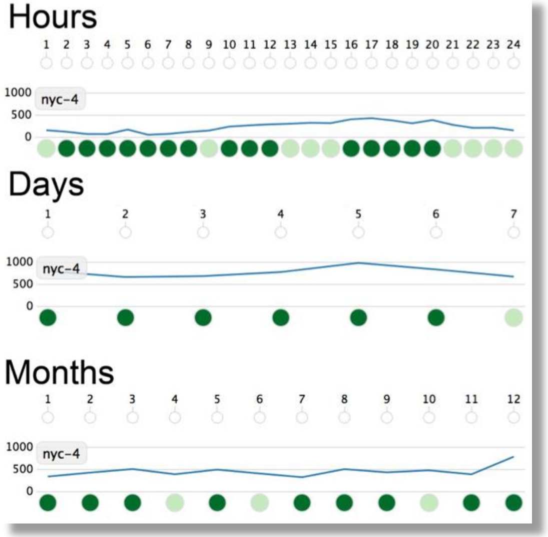
NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

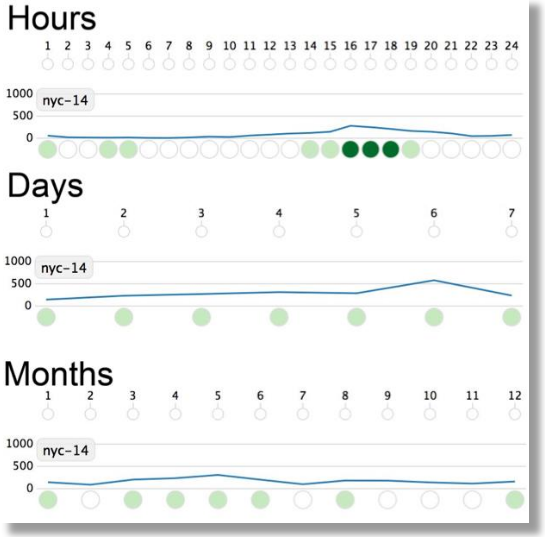
Quantify and compare "activity"

- Urban Pulse [Miranda et al., 2017]
  - Signature for different locations
  - Data oblivious
  - Rank and compare locations
  - Query similar locations

Rockefeller Center



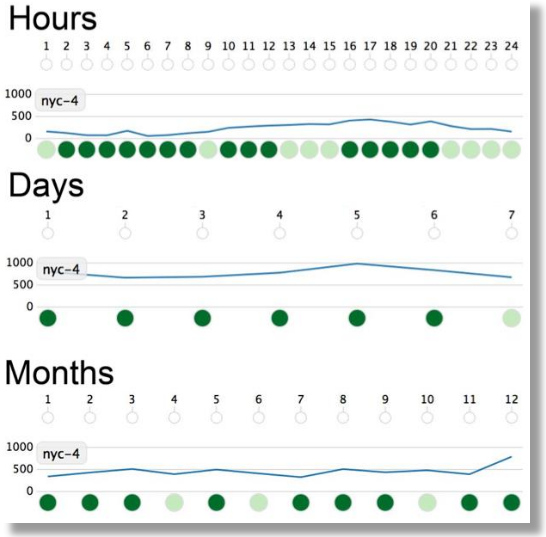
Union Square



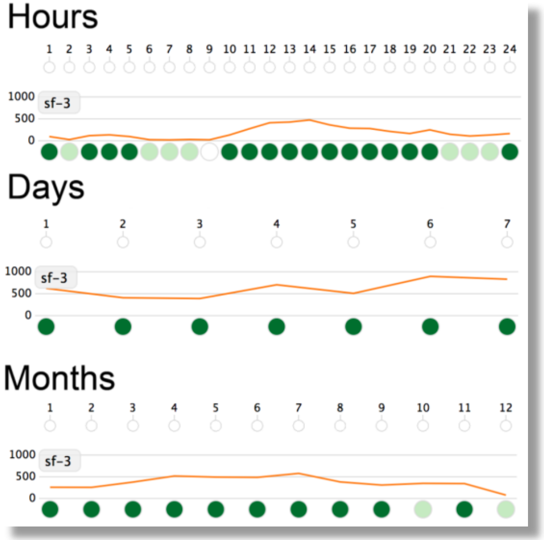
Quantify and compare "activity"

- Urban Pulse [Miranda et al., 2017]
  - Signature for different locations
  - Data oblivious
  - Rank and compare locations
  - Query similar locations

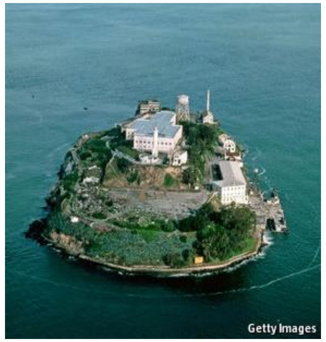
Rockefeller Center



San Francisco



Alcatraz

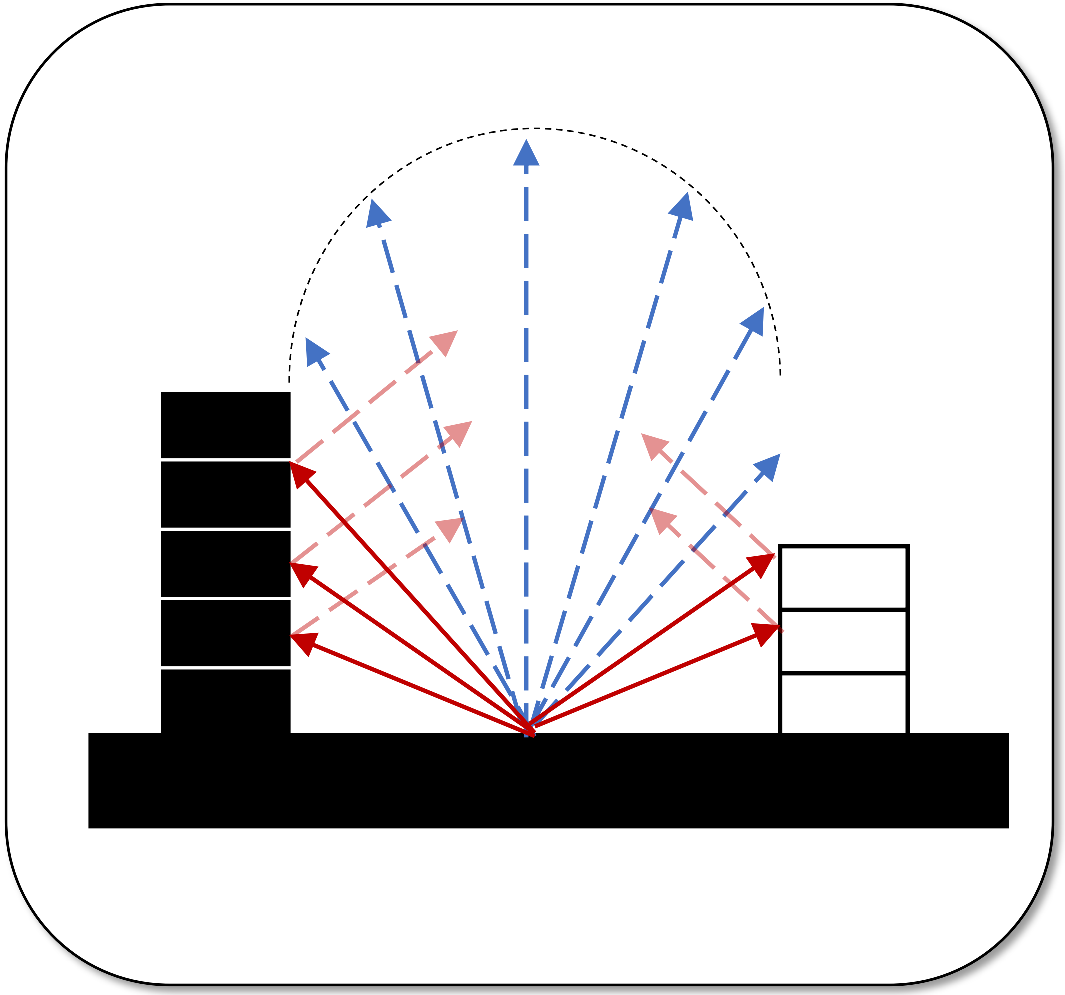


NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Analysis of sound propagation

- Potential approach
  - Make use of highly detailed building models available in NYC
  - Use ray tracing to propagate sound over time
- Initial technology already in place [Miranda, Doraiswamy et al., 2018]
  - Interactively compute shadow accumulation over time
  - Makes use of accurate 3D geometry
  - Uses GPU for efficient ray propagation



NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

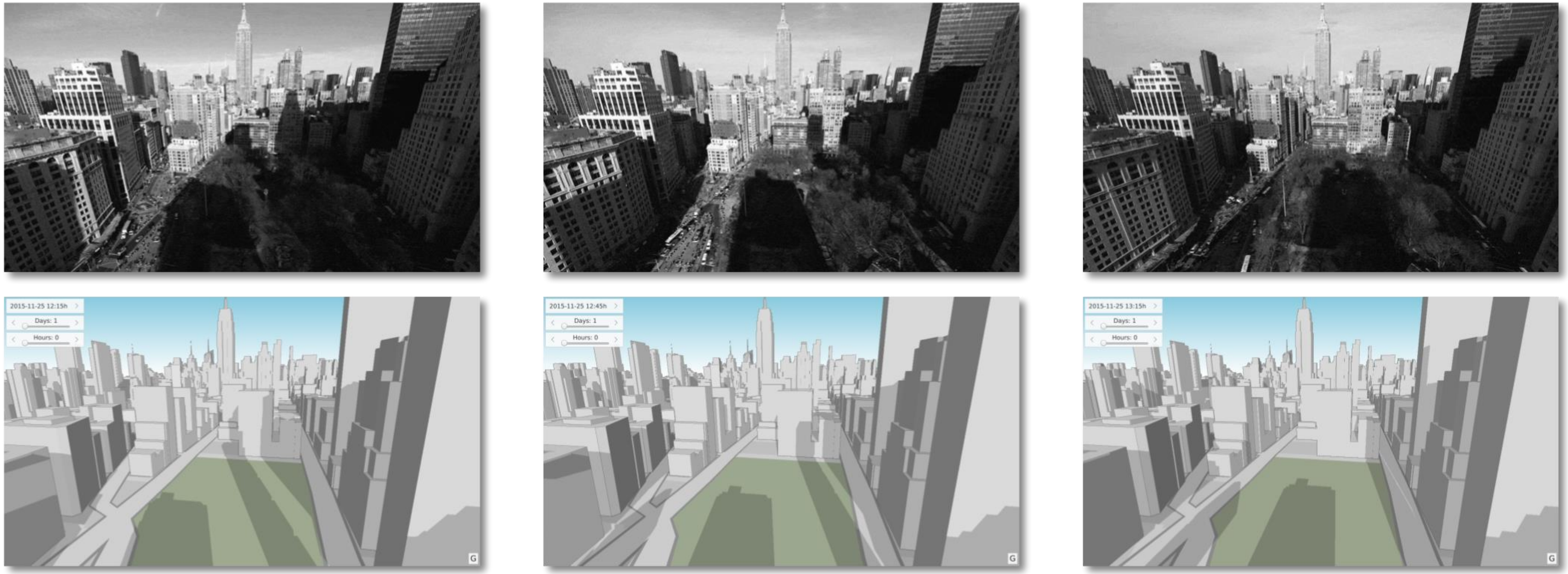
Quantifying shadow

- Shadow accumulation

[Mapping the Shadows of New York City - The New York Times]

Quantifying shadow

- Shadow accumulation

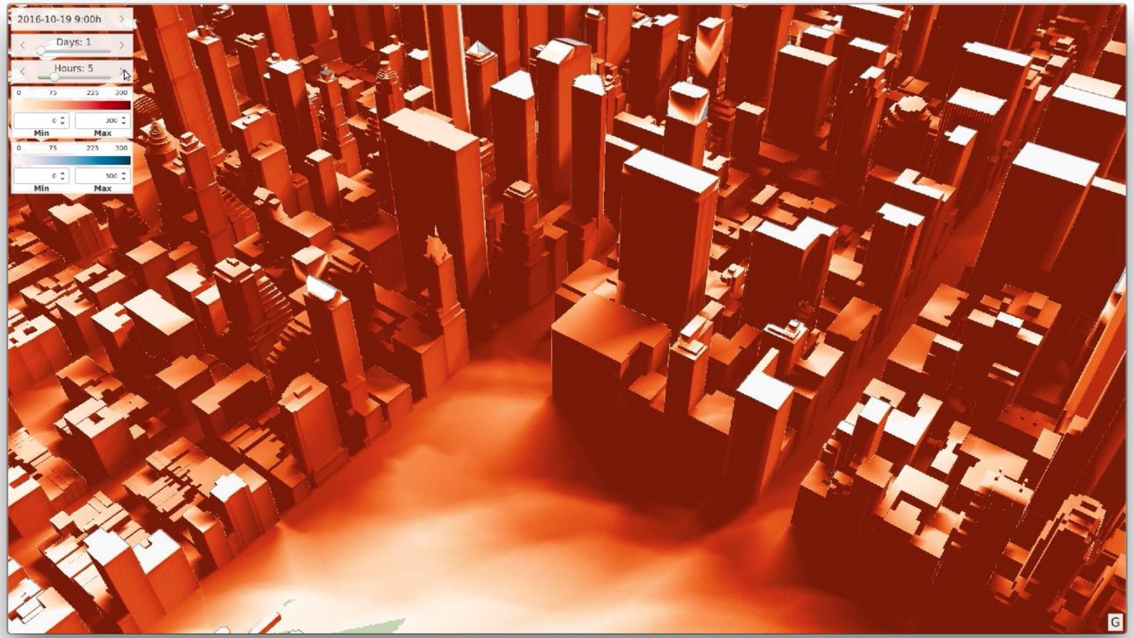


NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

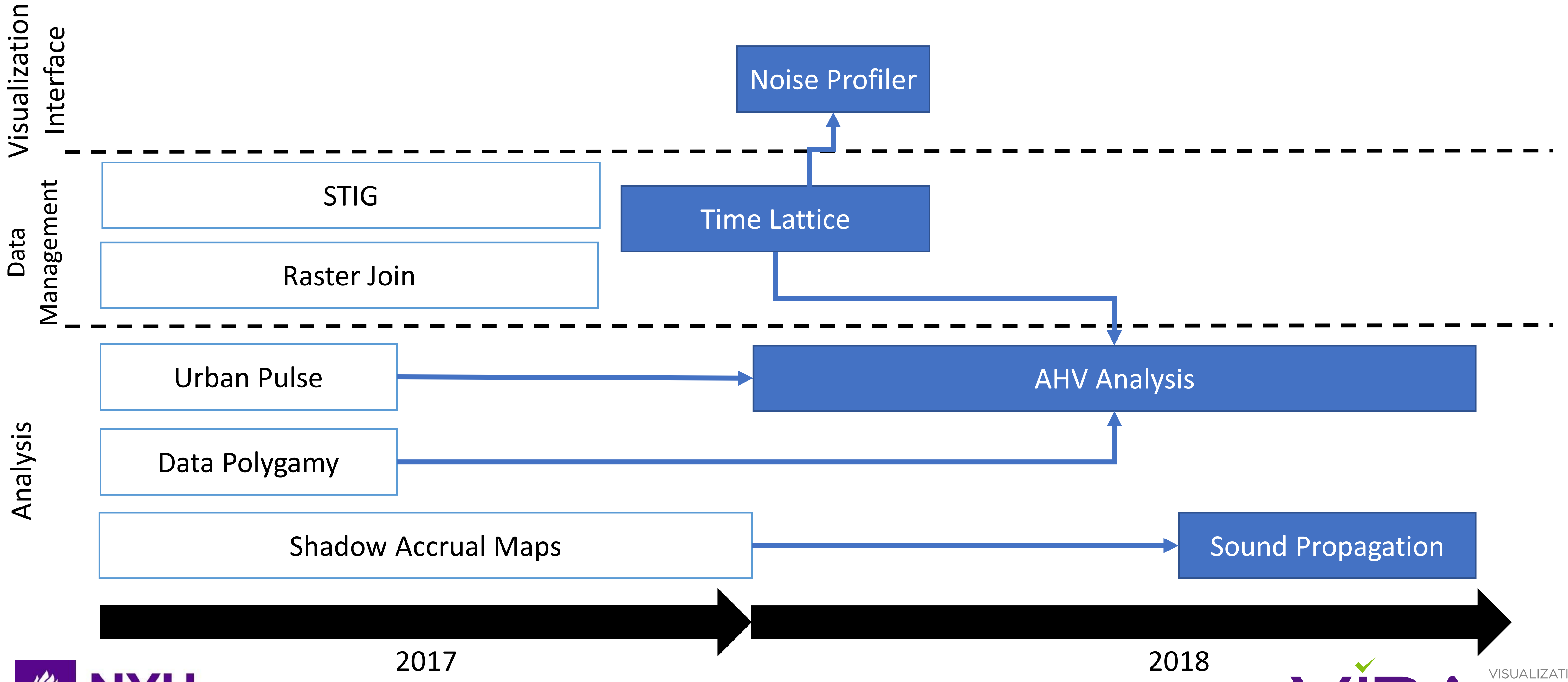
Quantifying shadow

- Shadow accumulation
  - Uses ray tracing to accumulate shadow over time
  - Allows for interactivity
  - Analysis of shadow impact from proposed buildings on public spaces



[Miranda, Doraiswamy et al., 2018]

Recap


Visualization
Interface
Data
Management
Analysis
Noise Profiler
STIG
Time Lattice
Raster Join
Urban Pulse
AHV Analysis
Data Polygamy
Shadow Accrual Maps
Sound Propagation
2017
2018

Outcomes

- Papers:

Published:

IEEE SciVis

TVCG

PVLDB

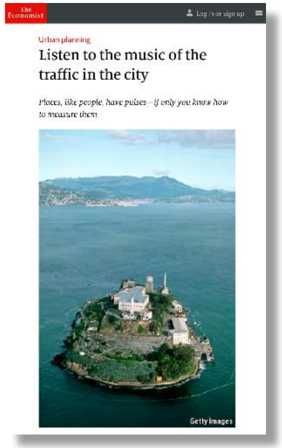
Submitted:

Eurovis

- Open source projects:
  - Raster Join: github.com/ViDA-NYU/raster-join
  - Urban Pulse: github.com/ViDA-NYU/urban-pulse
  - Time Lattice: soon

- Media coverage

New York Times



The Economist

Architecture Digest

Curbed

Thank you!

NYU

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER