Scientific Discovery Environments: A $_{\text{bias}}$ View Towards "The Next Generation"

Claudio T. Silva

Tandon School of Engineering

Center for Data Science

Center for Urban Science + Progress

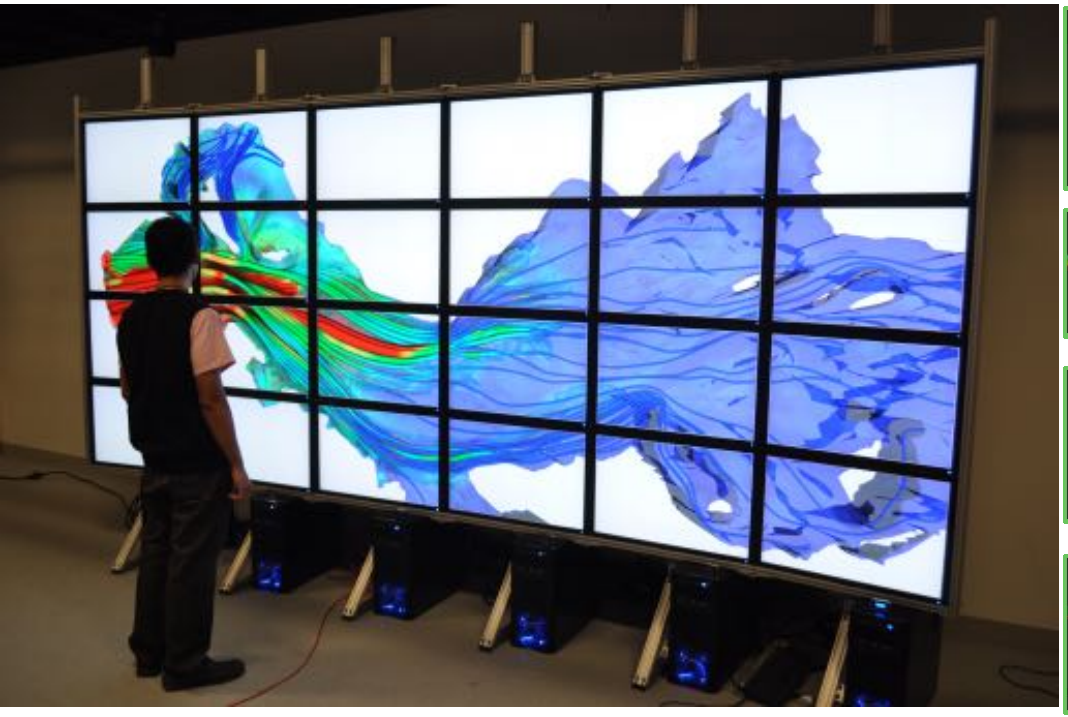
Courant Institute for Mathematical Sciences

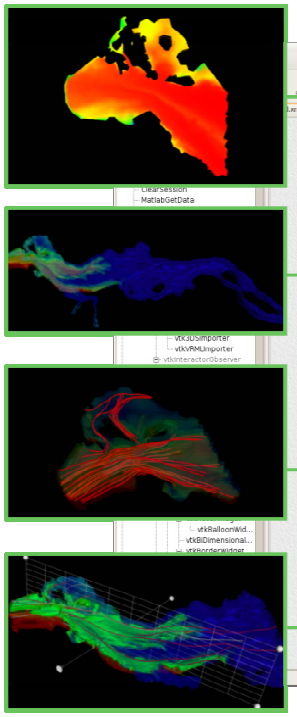
New York University

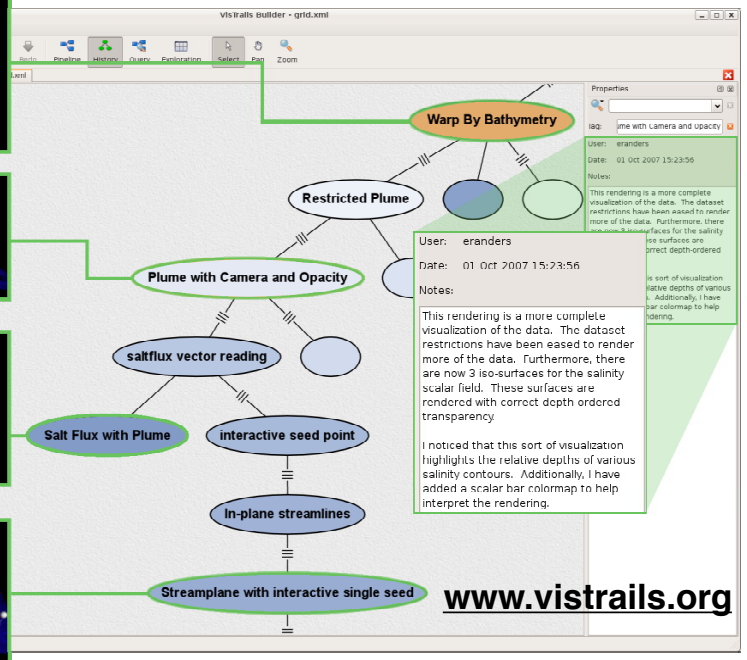
Funded by Moore and Sloan Foundations,
DARPA, NSF, NASA, DOE, MLB.com, AT&T, NVIDIA, and IBM

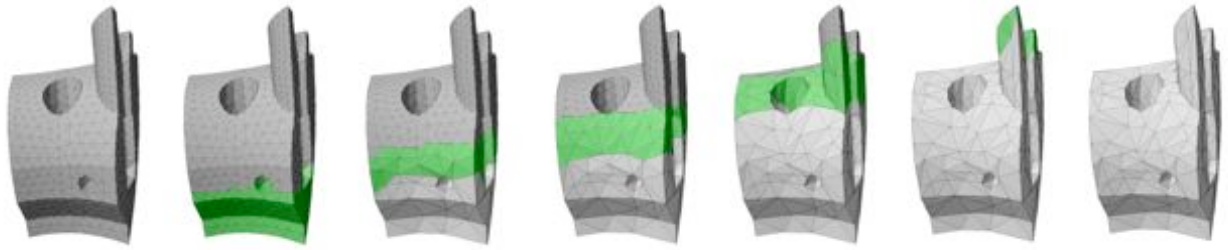
NYU TANDON SCHOOL OF ENGINEERING

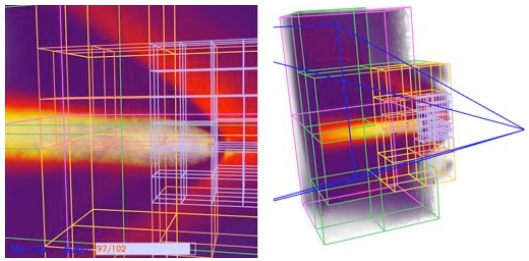
VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER



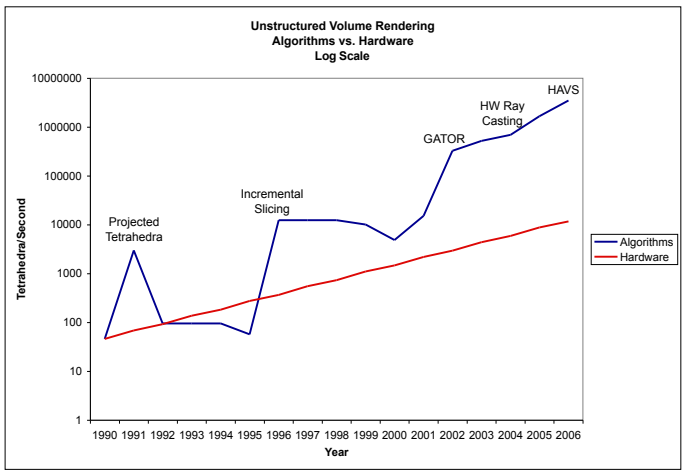


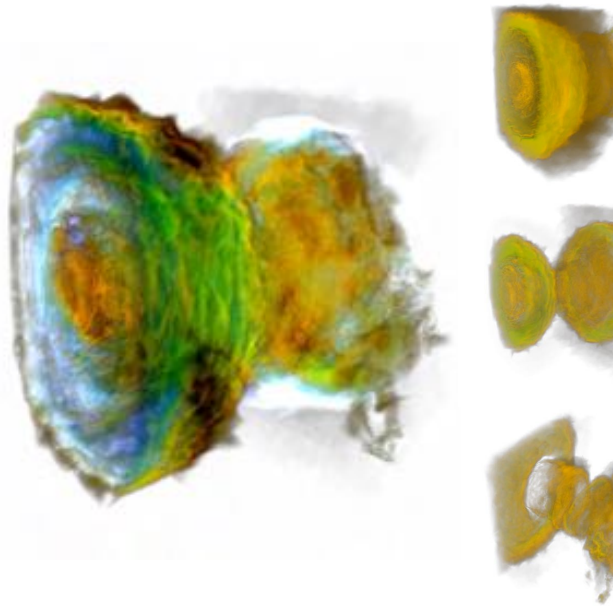

vistrails Builder - grid.xml
Warp By Bathymetry
Restricted Plume
Plume with Camera and Opacity
saltflux vector reading
Salt Flux with Plume
interactive seed point
In-plane streamlines
Streamplane with interactive single seed
User: eranders
Date: 01 Oct 2007 15:23:56
Notes:
This rendering is a more complete visualization of the data. The dataset restrictions have been eased to render more of the data. Furthermore, there are now 3 iso-surfaces for the salinity scalar field. These surfaces are rendered with correct depth ordered transparency.
I noticed that this sort of visualization highlights the relative depths of various salinity contours. Additionally, I have added a scalar bar colormap to help interpret the rendering.
www.vistrails.org






SDK 10 Update: New Stencil Routed K-Buffer Sample
DEVELOPER ZONE
SDK 10 Update: New Stencil Routed K-Buffer Sample
Could Your Game Run 35% Faster?
NVPerfHUD 4


Unstructured Volume Rendering
Algorithms vs. Hardware
Log Scale
Tetrahedra/Second
Year
Projected Tetrahedra
Incremental Slicing
GATOR
HW Ray Casting
HAVS
Algorithms
Hardware



Sports Data: MLB.com Statcast

Big Urban Data: Understanding Cities

Infrastructure

Condition, operations

Environment

Meteorology, pollution, noise, flora, fauna

People

Relationships, economic activities, health, nutrition, opinions, …

- City components interact in complex ways
- Need to look at the city *data exhaust* to understand these interactions
- Processes occur over time and space

twitter

DATA.GOV
EMPOWERING PEOPLE

YouTube
Broadcast Yourself

NYC OpenData

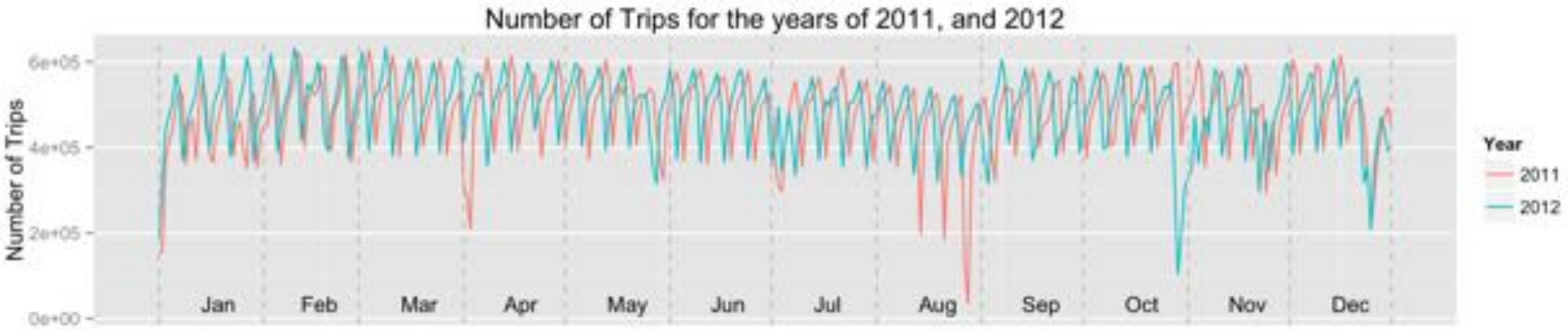
BRASIL Acesso à informação
Dados Abertos
GOVERNO FEDERAL

facebook

data.gov.in
Open Government Data Platform India
सत्यमेव जयते

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
data.gouv.fr

Exploring Urban Data: NYC Taxis



- Taxis are *sensors* that can provide unprecedented insight into city life: economic activity, human behavior, mobility patterns, …

  "How the taxi fleet activity varies during weekdays?"

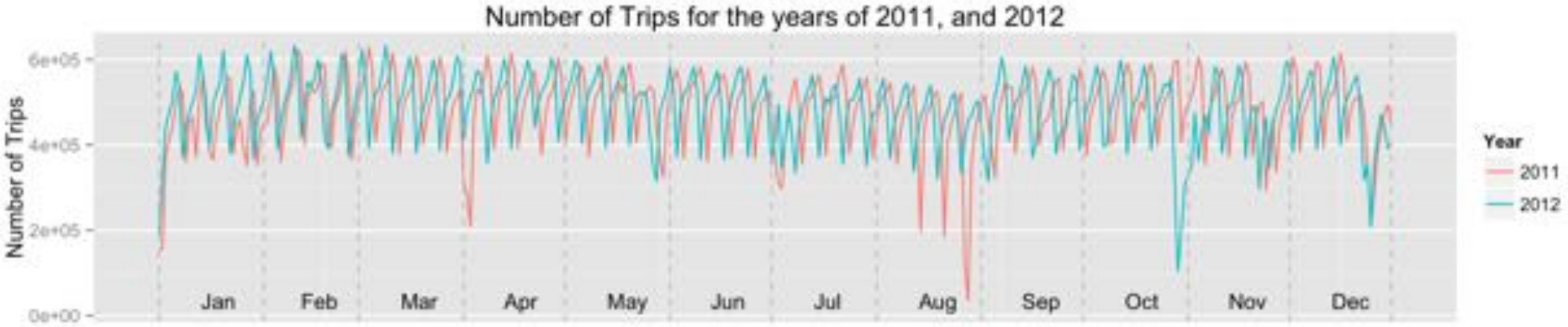
  "What is the average trip time from Midtown to the airports during weekdays?''

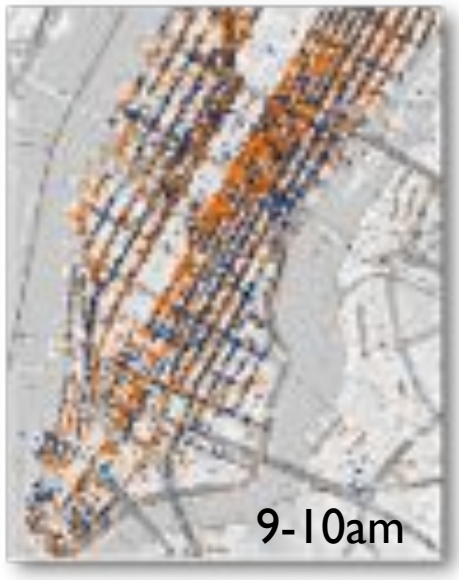
  "How was activity in Midtown affected during a presidential visit?''

  "How did the movement patterns change during Sandy?"

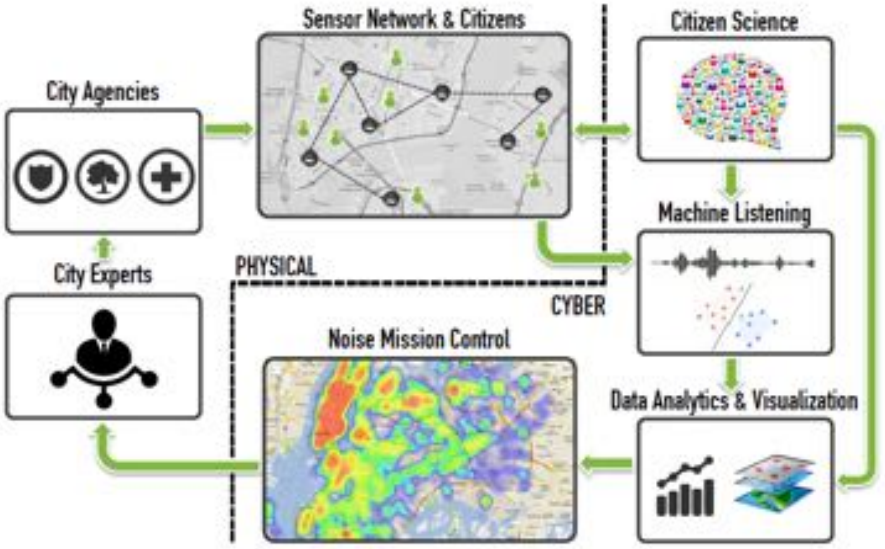
  "Where are the popular night spots?"

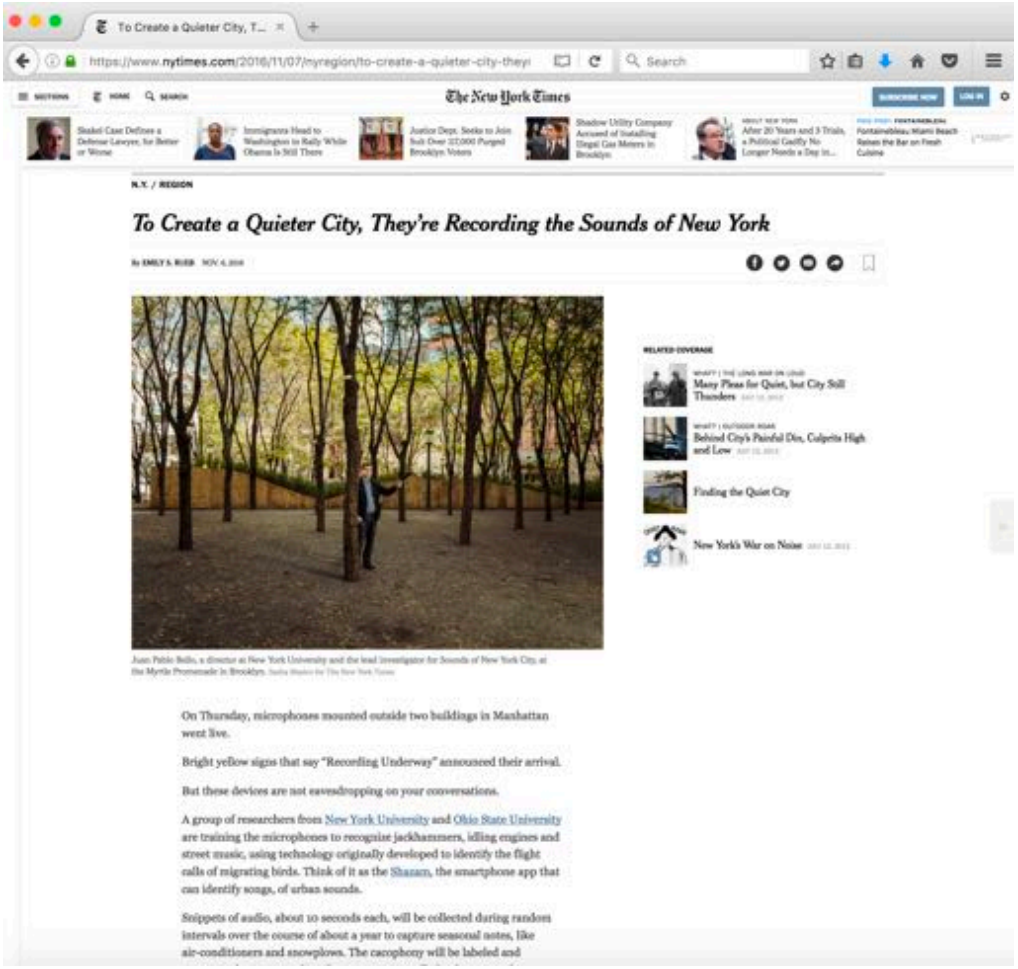
Exploring Urban Data: NYC Taxis





Sounds of New York City





To Create a Quieter City, T...

https://www.nytimes.com/2016/11/07/nyregion/to-create-a-quieter-city-they

The New York Times

N.Y. / REGION

To Create a Quieter City, They're Recording the Sounds of New York

Many Pleas for Quiet, but City Still Thunders

Behind City's Painful Din, Culprits High and Low

Finding the Quiet City

New York's War on Noise

On Thursday, microphones mounted outside two buildings in Manhattan went live.

Bright yellow signs that say "Recording Underway" announced their arrival.

But these devices are not eavesdropping on your conversations.

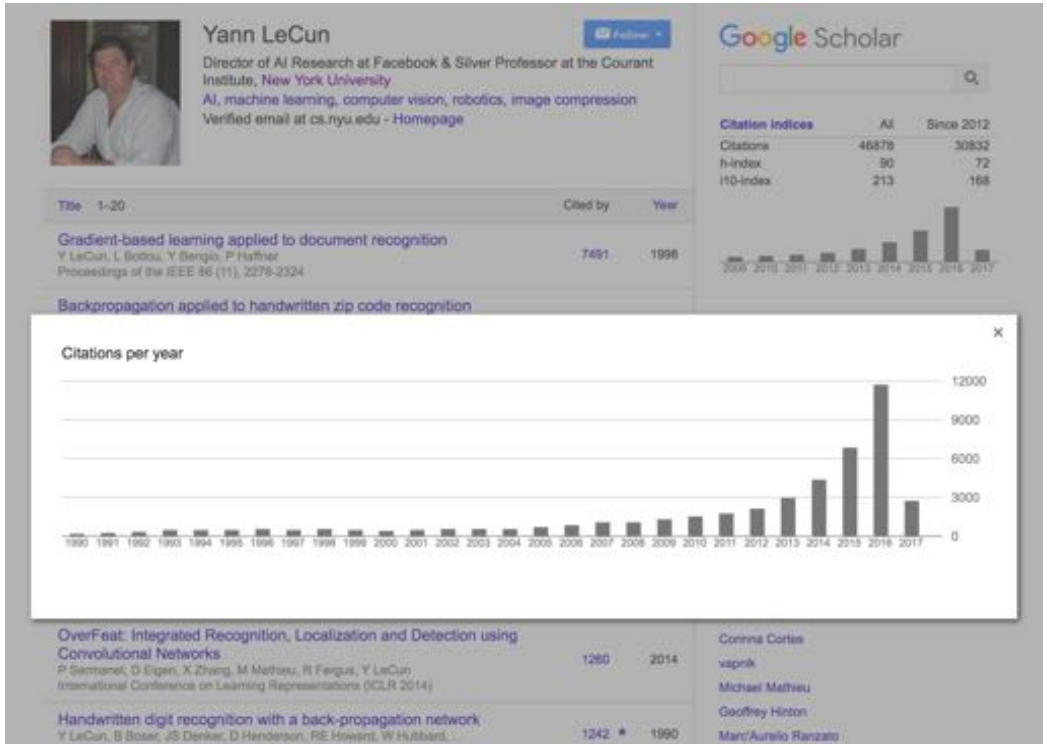
A group of researchers from New York University and Ohio State University are training the microphones to recognize jackhammers, idling engines and street music, using technology originally developed to identify the flight calls of migrating birds. Think of it as the Shazam, the smartphone app that can identify songs, of urban sounds.

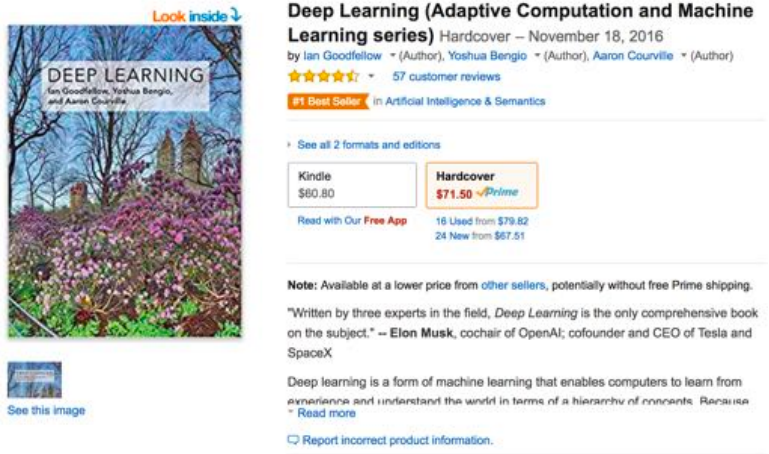
Snippets of audio, about 10 seconds each, will be collected during random intervals over the course of about a year to capture seasonal notes, like air-conditioners and snowplows. The cacophony will be labeled and


PAPAYA DOG
OPEN

Major Trends

- AI
  - Deep learning
  - Machine learning
  - Natural language
  - Automatic analysis
  - Anomaly detection
- Data management
  - Linking analysis and data
  - Provenance
  - Collaborative support
  - Lots of simultaneous datasets
  - Lots of visualizations
  - Lots of images
- GUI
  - Desktop
  - Touch-enabled interfaces
  - Large Displays
  - VR/AR
  - General versus problem specific
- System support/architecture
  - Cloud environments
  - Interactive support/programming
  - Progressive Visualization
  - Client-server support
  - Parallelism

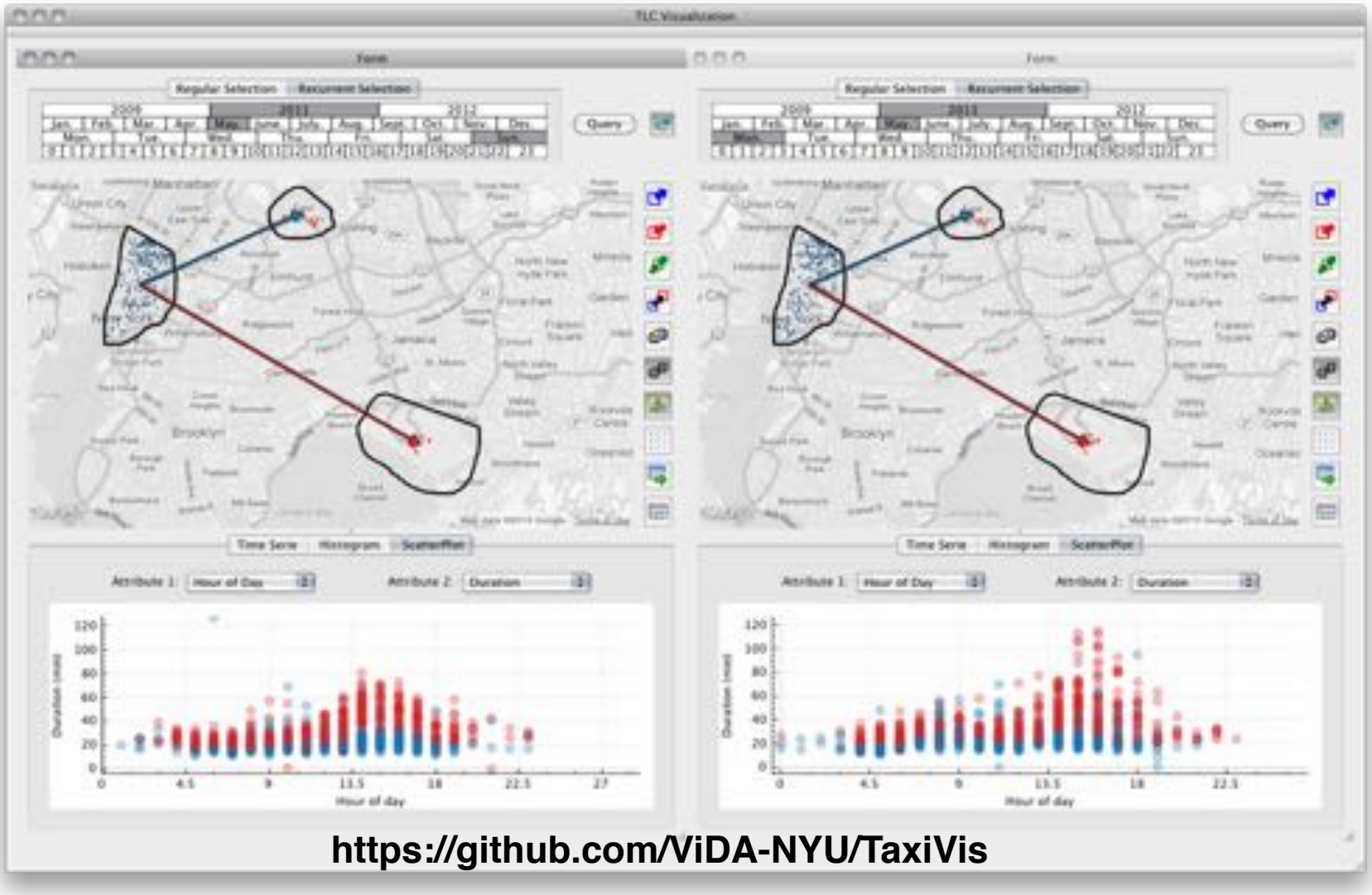




NYU TANDON SCHOOL OF ENGINEERING

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Graphical User Interfaces
General versus problem specific



https://github.com/ViDA-NYU/TaxiVis

Ferreira et al, 2013

Provenance

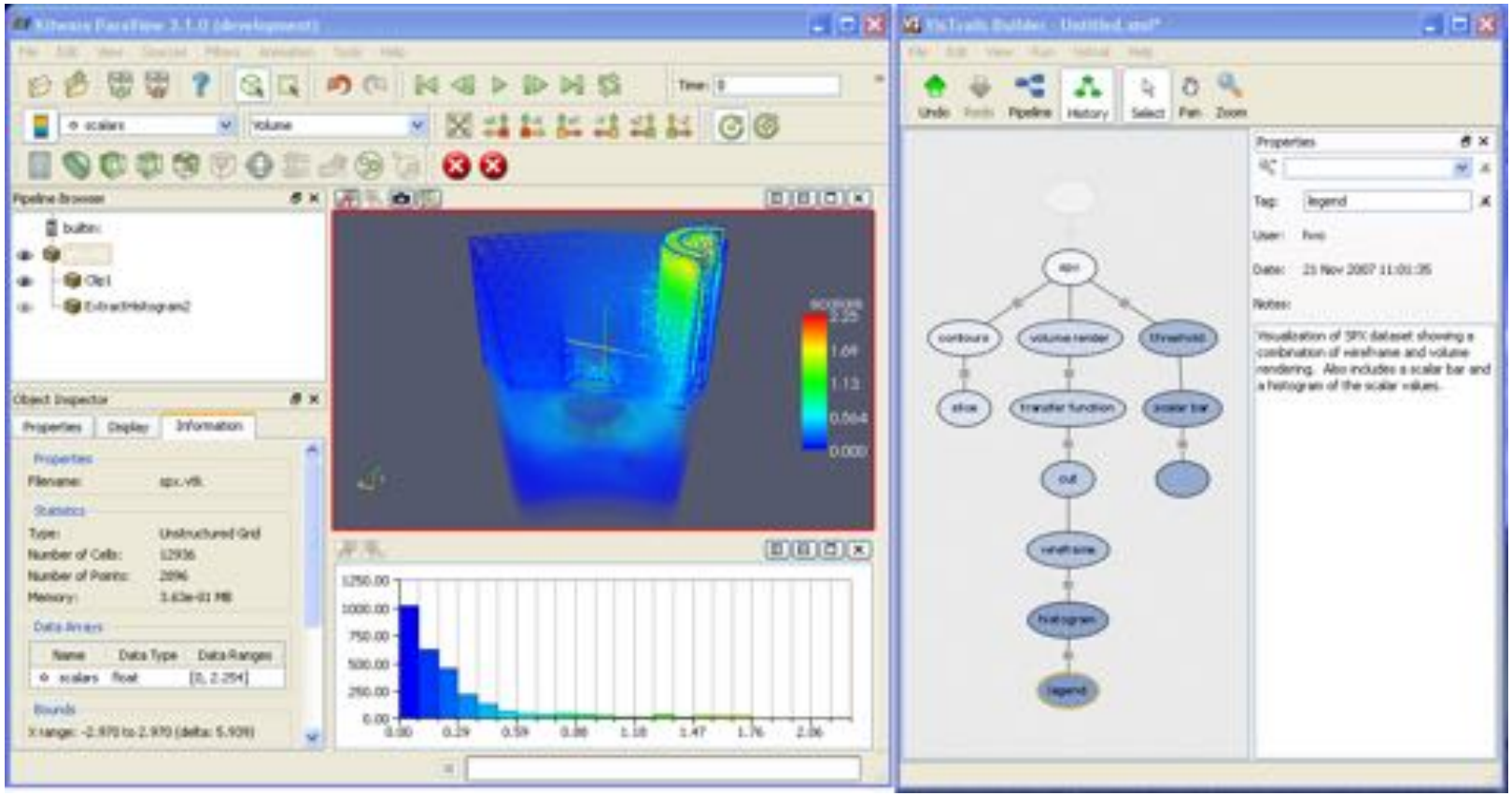


Fig. 2. A screenshot of ParaView (left) with the provenance captured by VisTrails and displayed as a version tree in a separate window (right). This preliminary prototype taps into ParaView undo/redo mechanism to capture the exploration process.

Callahan et al, 2008

Provenance

VisTrails Plugin for ParaView

Data management / UrbanGIS

Urbane: A 3D Framework to Support Data Driven Decision Making in Urban Development

Nivan Ferreira* New York University
Marcos Lage† Universidade Federal Fluminense
Harish Doraiswamy‡ New York University
Huy Vo§ New York University
Luc Wilson¶ Heidi Werner¶ Muchan Park¶ Kohn Pedersen Fox Associates PC
Cláudio Silva‖ New York University

Figure 1: Urbane provides architects, developers, and planners with a new, data and analysis rich way of reading the city with the goal of improving decision making in urban development. Users can explore properties of neighborhoods and buildings using the data exploration view to identify underdeveloped sites for potential development. Then, using the visual interface together with the map view, they can simulate the impact of such development. For example, the views of the freedom tower (highlighted in green) from the buildings highlighted in red would be adversely impacted (positively impacted buildings are highlighted in blue) if the new constructions (colored yellow) are built. The supplemental video shows the different features and visualizations supported by Urbane.

ABSTRACT

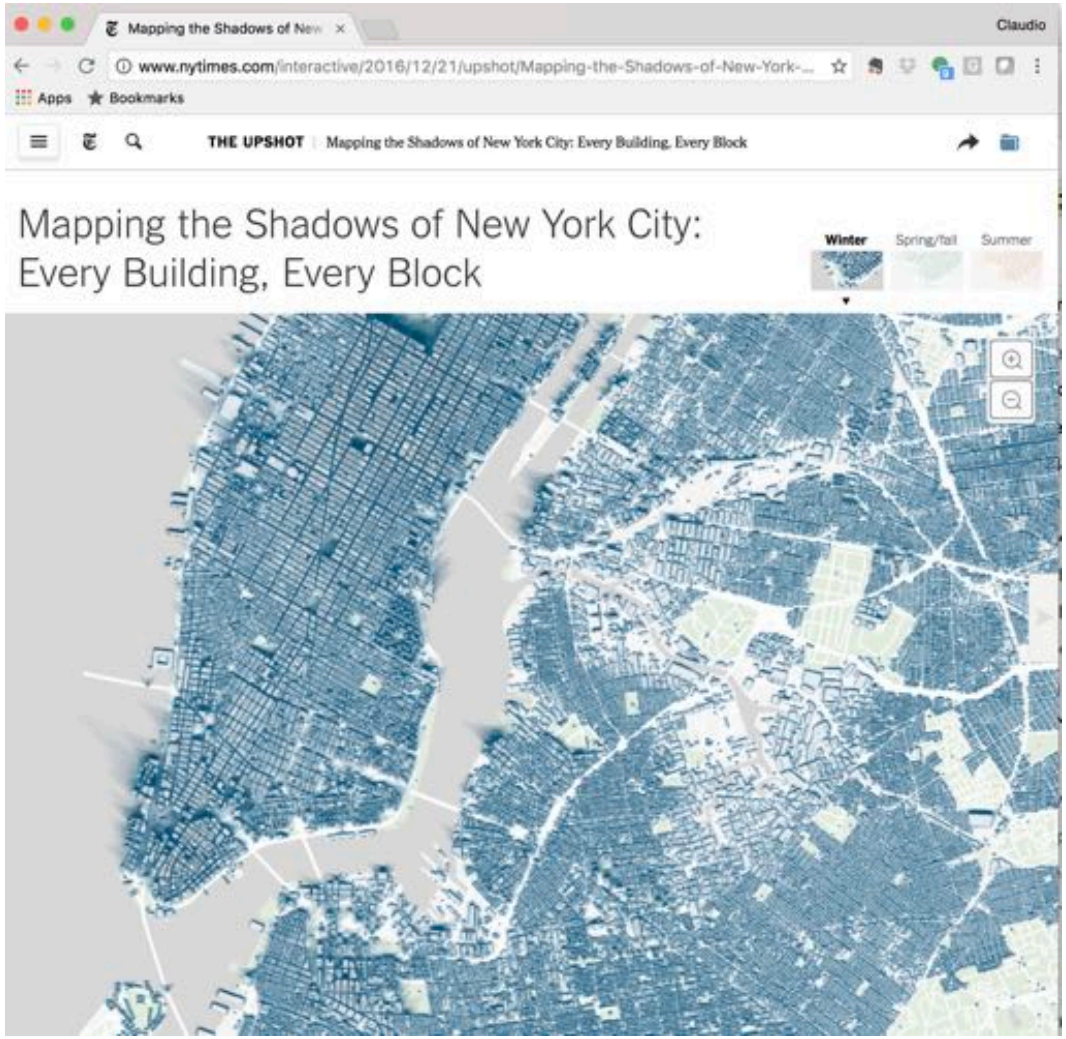
Architects working with developers and city planners typically rely on experience, precedent and data analyzed in isolation when making decisions that impact the character of a city. These decisions are critical in enabling vibrant, sustainable environments but must also negotiate a range of complex political and social forces. This requires those shaping the built environment to balance maximizing the value of a new development with its impact on the character of a neighborhood. As a result architects are focused on two issues throughout the decision making process: a) what defines the character of a neighborhood? and b) how will a new development change its neighborhood? In the first, character can be influenced by a variety of factors and understanding the interplay between diverse data sets is crucial; including safety, transportation access, school quality and access to entertainment. In the second, the impact of a new development is measured, for example, by how it impacts the view from the buildings that surround it. In this paper, we work in collaboration with architects to design Urbane, a 3-dimensional multi-resolution framework that enables a data-driven approach for decision making in the design of new urban development. This is accomplished by integrating multiple data layers and impact analysis techniques facilitating architects to explore and assess the effect of these attributes on the character and value of a neighborhood. Several of these data layers, as well as impact analysis, involve working in 3-dimensions and operating in real time. Efficient computation and visualization is accomplished through the use of techniques from computer graphics. We demonstrate the effectiveness of Urbane through a case study of development in Manhattan depicting how a data-driven understanding of the value and impact of speculative buildings can benefit the design-development process between architects, planners and developers.

Keywords: Urban data analysis; GIS; impact analysis; visual analytics; architecture; city development

1 INTRODUCTION

Why do two neighborhoods feel similar? Or different? Why does a new building change the quality of a neighborhood and another doesn't? While the experience of a city is inherently subjective, the characteristics that shape the quality of it are not. These characteristics can be difficult to obtain, measure or analyze by those shaping the future of a city. Architects working with developers and city planners typically rely on experience, precedent and data analyzed in isolation when making decisions that impact the character of a city. These decisions, while being critical in enabling vibrant and sustainable environments, must also negotiate a range of complex political and social forces. This requires those shaping the built environment to balance maximizing the value of new development

*e-mail:nivan.ferreira@nyu.edu
†e-mail:mlage@ic.uff.br
‡e-mail:harishd@nyu.edu
§e-mail:huy.vo@nyu.edu
¶e-mail:{lwilson,hwerner,mpark}@kpf.com
‖e-mail:csilva@nyu.edu



Ferreira et al, 2015

NYU TANDON SCHOOL OF ENGINEERING

VIDA VISUALIZATION IMAGING AND DATA ANALYSIS CENTER

Data management / UrbanGIS

URBANE:

A 3D Framework to Support Data Driven Decision Making in Urban Developments

IEEE VAST 2015
Submission ID: 268

Harish Doraiswamy[1], Nivan Ferreira[1], Huy Vo[1], Claudio Silva[1], Marcos Lage[2], Muchan Park[3], Heidi Werner[3], Luc Wilson[3]

New York University[1], Universidade Federal Fluminense[2], Kohn Pederson Fox Associates PC[3]

Data management / Image Collections

ARIES: Enabling Visual Exploration and Organization of Art Image Collections

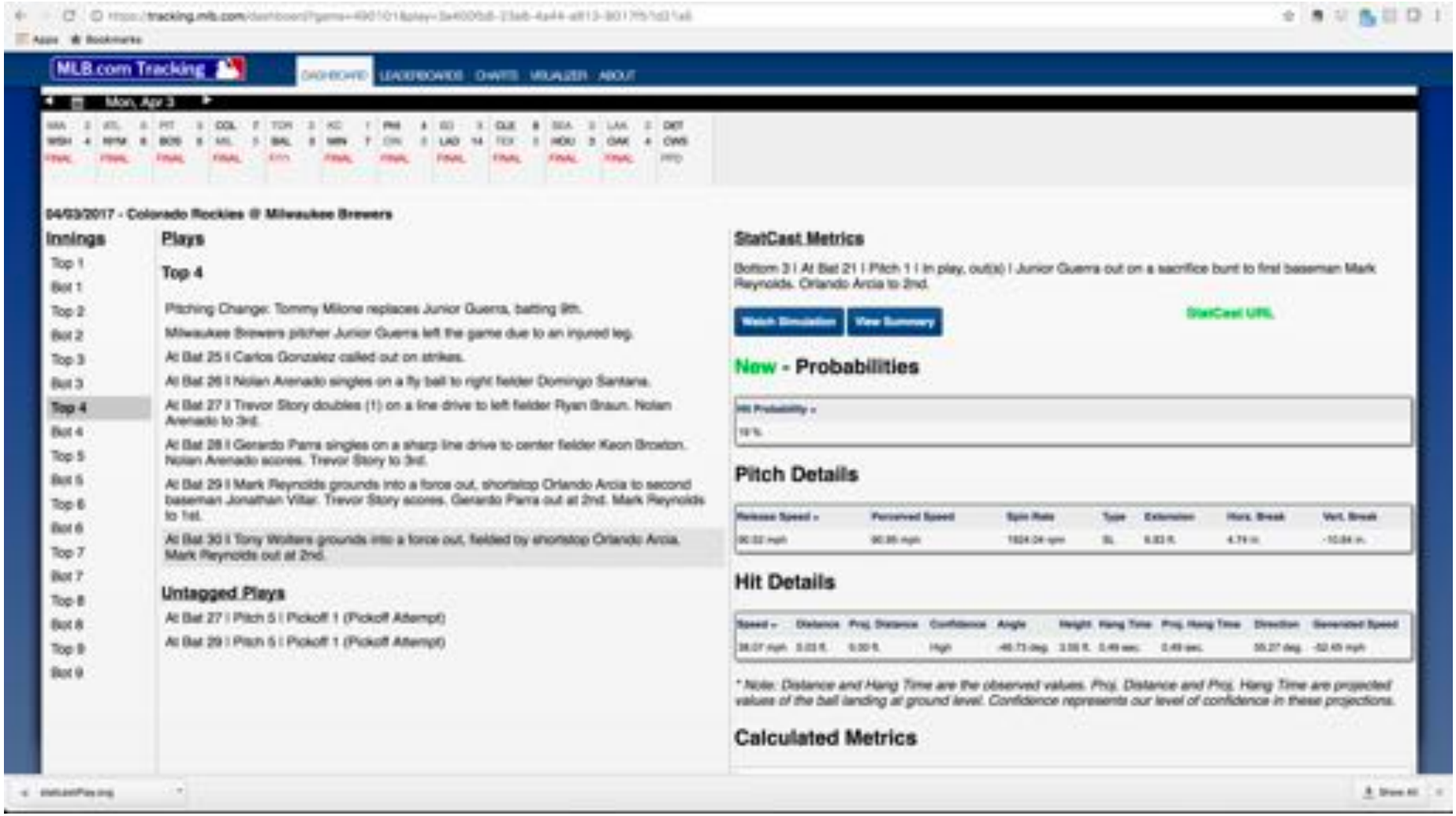
Lhaylla Crissaff, Louisa Ruby, Samantha Deutch, Luke DuBois, Jean-Daniel Fekete, *Senior Member, IEEE*, Juliana Freire, *Member, IEEE*, Cláudio T. Silva, *Fellow, IEEE*

Fig. 2. The ARIES interface includes a toolbar (a) and four views: image menu (b), metadata (c), lightbox canvas (d) and group menu (e). Image menu, metadata and group menu are retractable, enlarging the lightbox canvas. Works of art on the lightbox canvas are displayed in relative size.

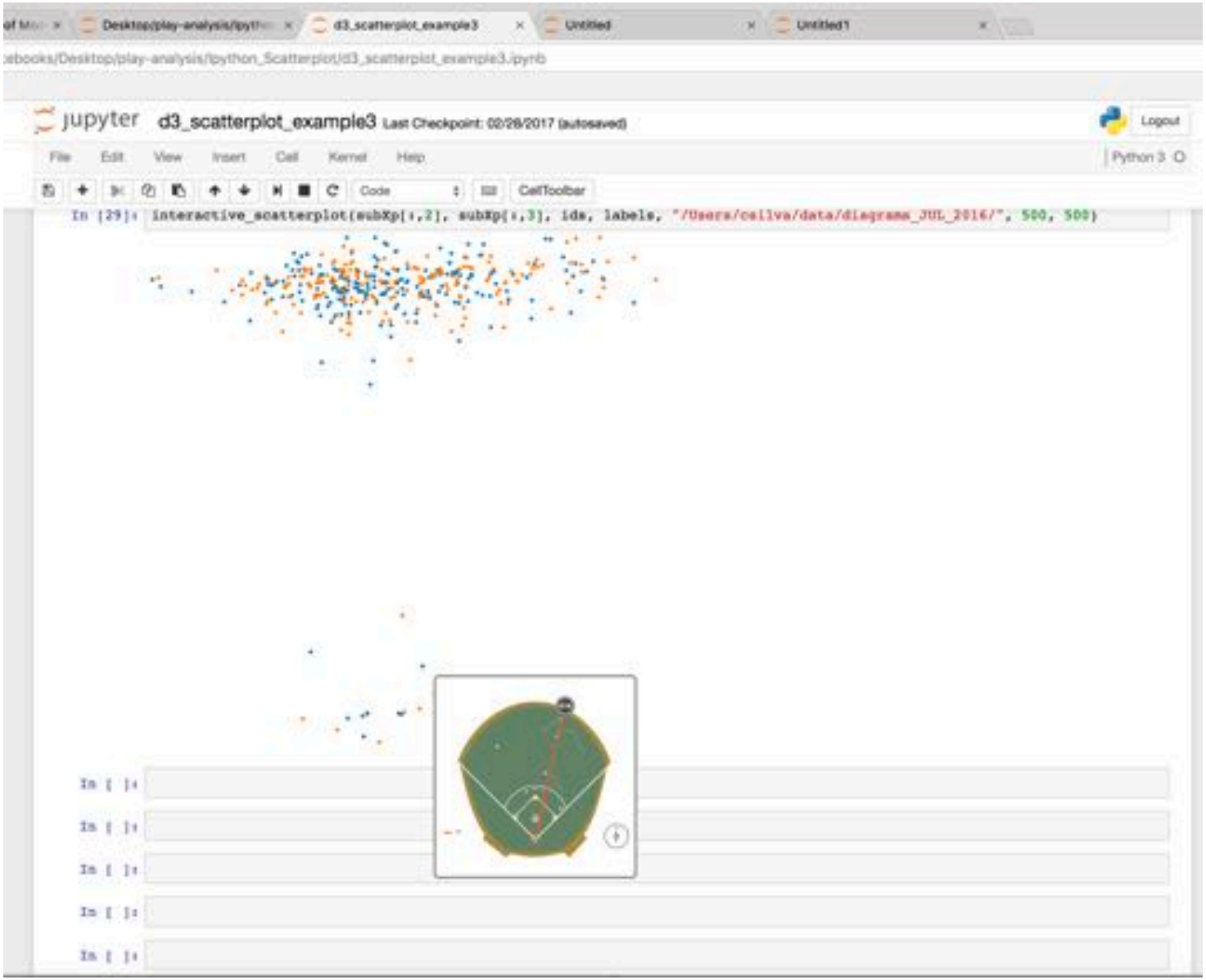
Data management / Image Collections

Data Management / Cloud

Live Demo: Statcast



System support / Interactive Programming



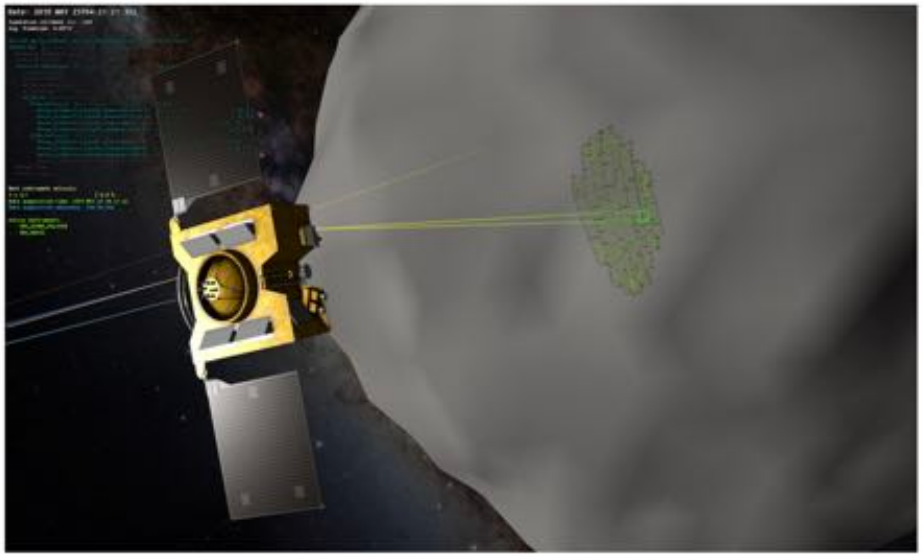
OpenSpace News Examples About Team Download Contact Development

OpenSpace is open source interactive data visualization software designed to visualize the entire known universe and portray our ongoing efforts to investigate the cosmos.

OpenSpace brings the latest techniques from data visualization research to the general public. OpenSpace supports interactive presentation of dynamic data from observations, simulations, and space mission planning and operations. OpenSpace works on multiple operating systems, with an extensible architecture powering high resolution tiled displays and planetarium domes, and makes use of the latest graphic card technologies for rapid data throughput. In addition, OpenSpace enables simultaneous connections across the globe, creating opportunity for shared experiences among audiences worldwide.

Osiris-REx Launch Event at AMNH

September 9, 2016 Kayla Nussbaum News



Today, NASA

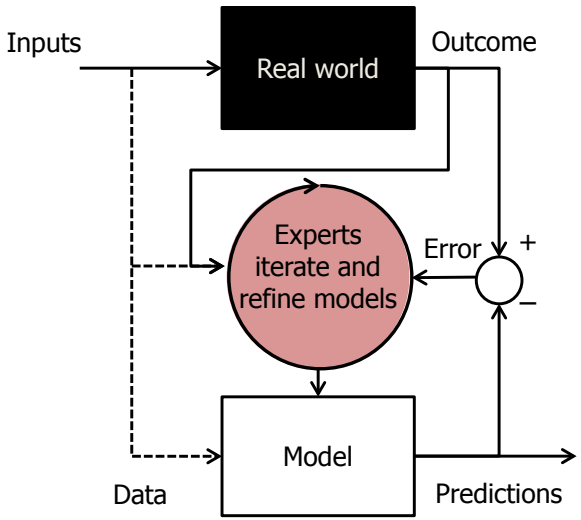
AI

DARPA D3M, Mr Wade Shen

http://www.darpa.mil/program/data-driven-discovery-of-models

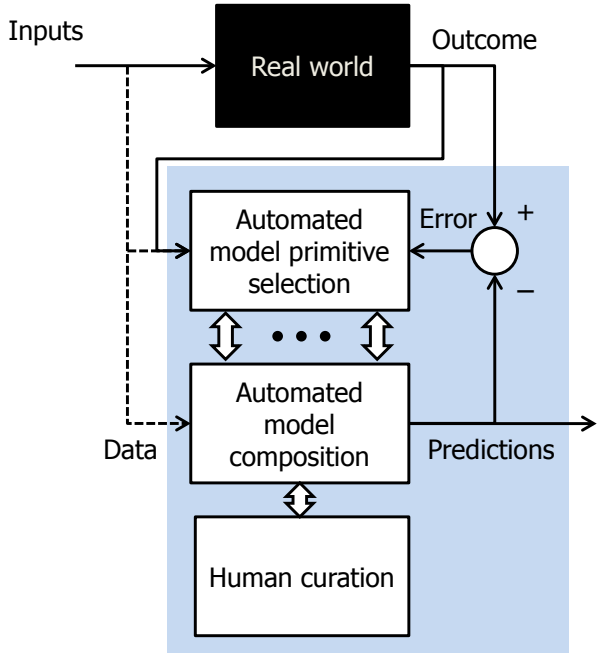
D^{3}M: Data-driven discovery of models

Today: Manual



- Model: representation of a real-world system
  - Examples
    - Inferring locations of images
    - Prediction of election outcomes
    - Estimation model for disease outbreaks
- Manual process: 10-1000s of person-years
- Teams of experts required to develop the model

Tomorrow: Automated



- Automatically select problem-specific model primitives
  - Extend the library of modeling primitives
- Automatically compose complex models from primitives
- Facilitate user interaction with composed models

AI / Natural Language

FlowSense: A Natural Language Interface for Visual Data Exploration with Data Flow

Bowen Yu and Cláudio T. Silva *Fellow, IEEE*

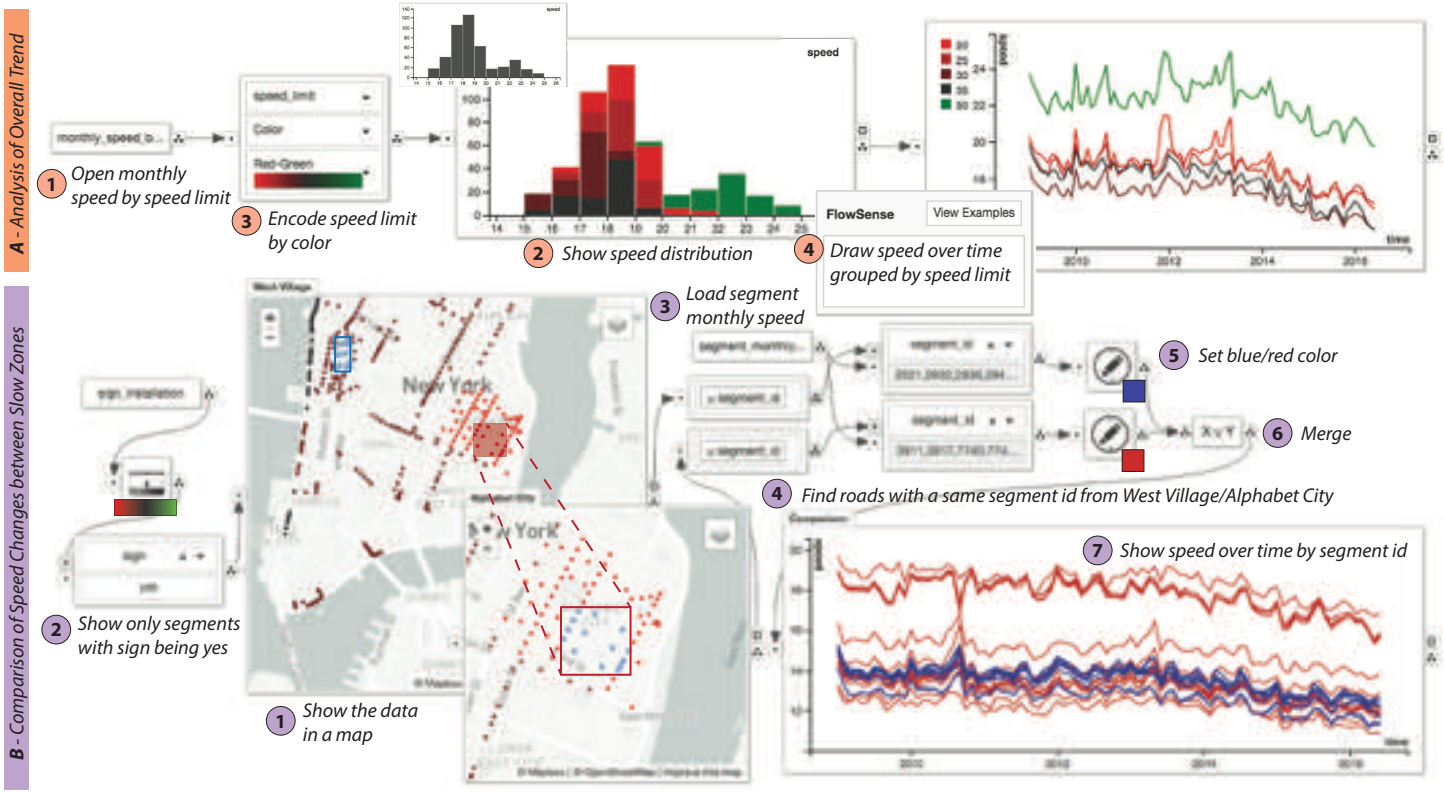
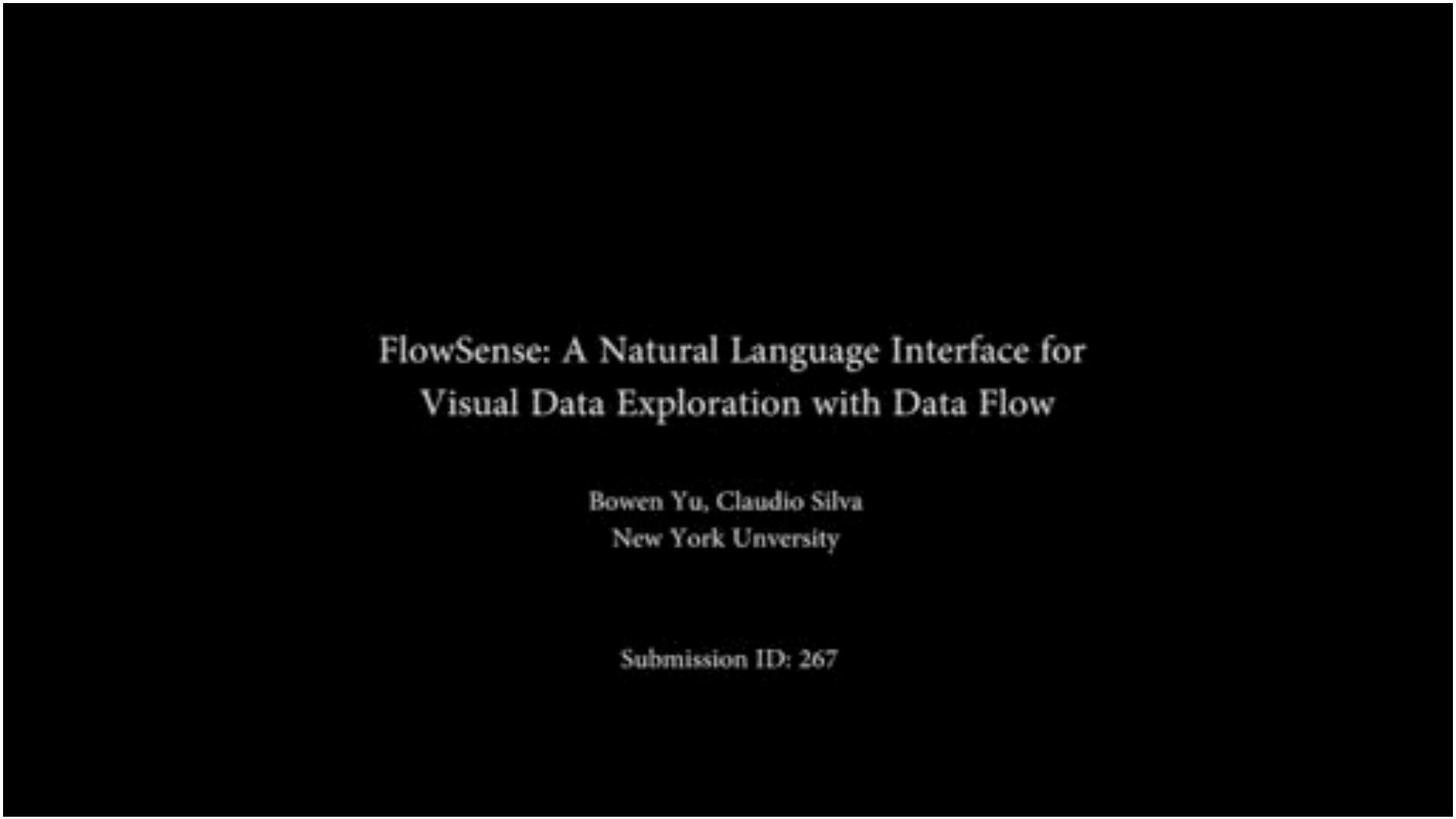


Fig. 1. Applying FlowSense to study the speed reduction in New York City. The important steps in the analysis and their natural language queries are shown in order. (A) FlowSense visualizes the overall speed reduction trend for streets of different speed limit. Step 4 shows the FlowSense dialog for typing queries. (B) A comparative study on the street speed changes between the West Village slow zone (blue) and the Alphabet City slow zone (red).

AI / Natural Language



AI / Features Over 1000s of Datasets

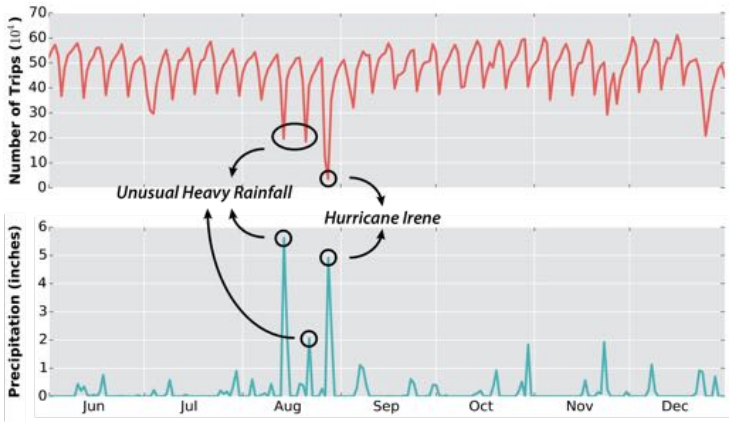


Fig. 2. Variation of the number of taxi trips in NYC over time (top) and its relationship with precipitation (bottom).

Chirigati et al, 2016

Chan et al, 2017

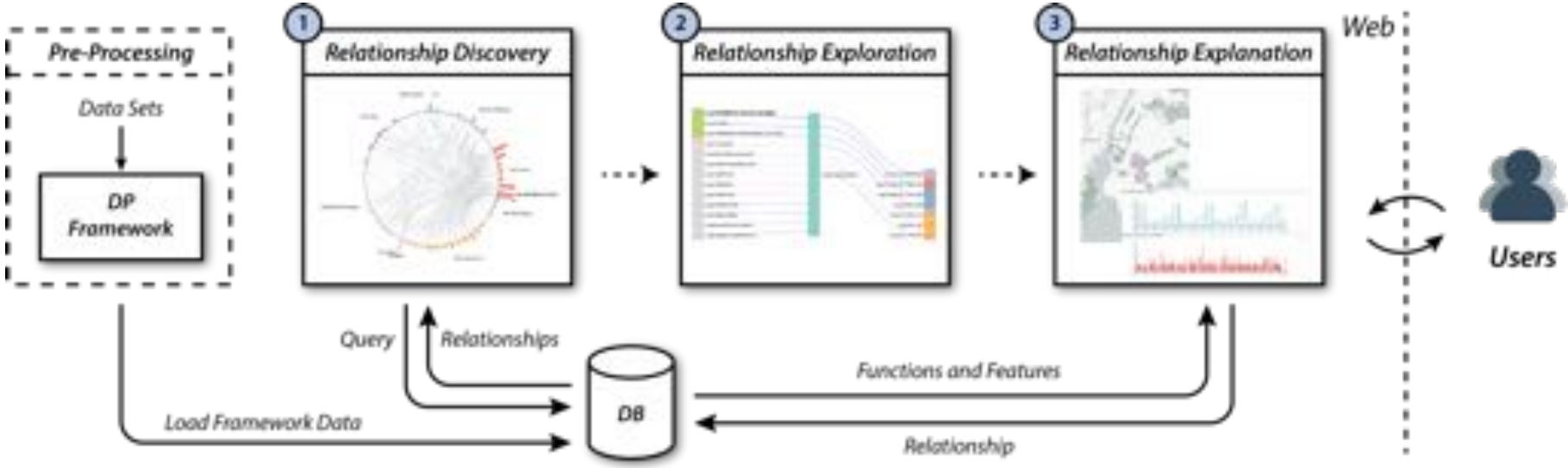
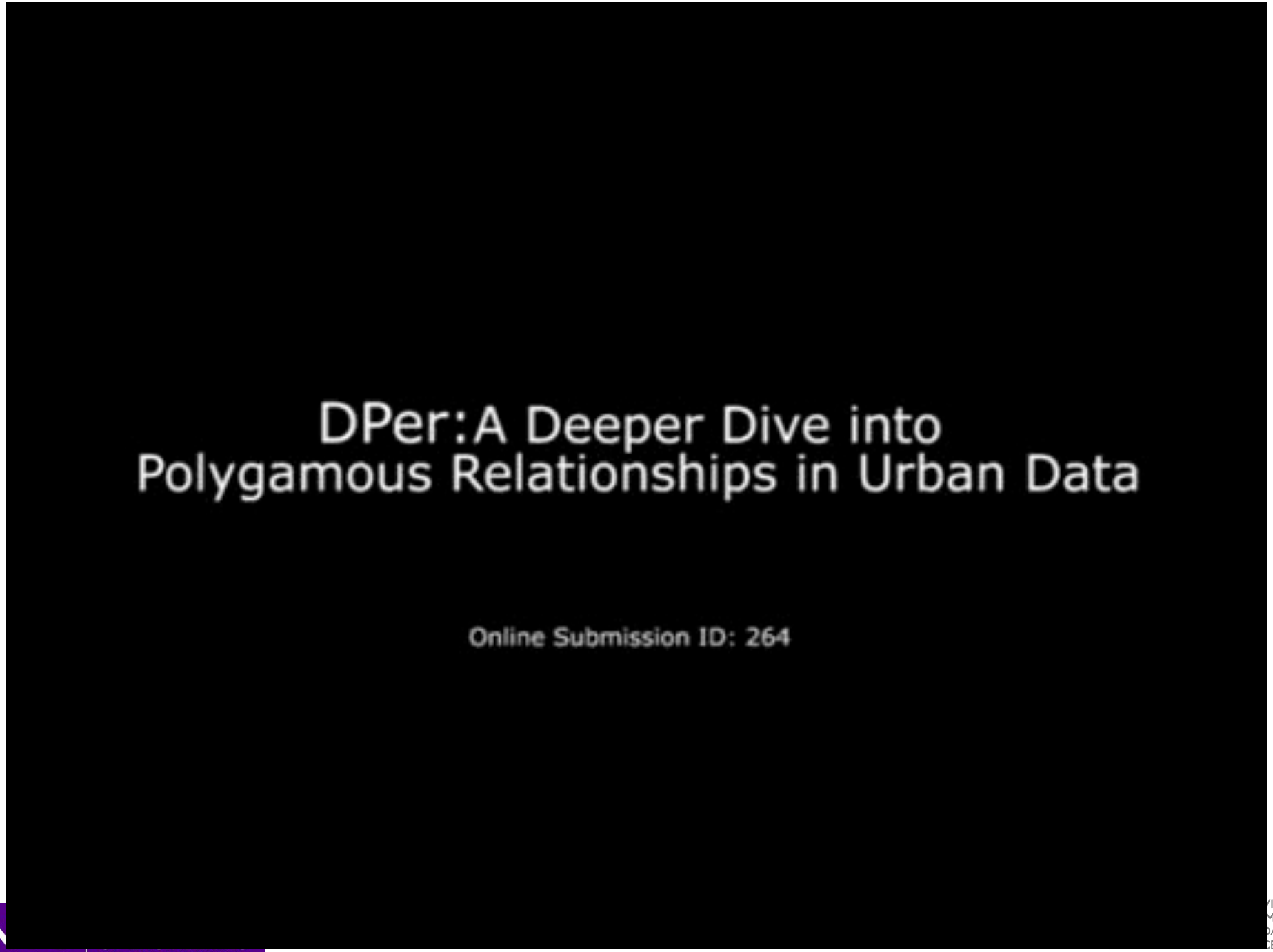


Fig. 4. DPER overview: the three components of the system correspond to the different stages of the data analysis pipeline. First, users query for interesting relationships (1); then, they can browse and filter the query results based on the relationships properties (2); finally, given a relationship, users can further inspect the data behind it to assess its validity (3).

AI / Features Over 1000s of Datasets



Major trends - Summary

Artificial Intelligence
- Deep learning

Data management

User Interfaces

System architecture
- Cloud environments
- Interactive support/programming

FOCUS ON THE USER!
—
DO NOT BE AFRAID TO GO OFF THE BEATEN PATH,
DEVELOP NEW SYSTEMS,
IDEAS ARE MORE IMPORTANT THAN CODE,
AND TECHNOLOGY MOVES VERY FAST
—
TRY NEW IDEAS!
RELEASE AS OPEN SOURCE!

Thank you!

csilva@nyu.edu

Urban Data Analysis @ NYU VIDA

Urbane: A 3D Framework to Support Data Driven Decision Making in Urban Development

A GPU-Based Index to Support Interactive Spatio-Temporal Queries over Historical Data

Harish Doraiswamy* Huy T. Vo† Cláudio Silva* Juliana Freire*
*New York University †The City College of the City University of New York
{harishd, huy.vo, csilva, juliana.freire}@nyu.edu

Topology-based Catalogue Exploration Framework for Identifying View-Enhanced Tower Designs

A Scalable Approach for Data-Driven Taxi Ride-Sharing Simulation

Masayo Ota[1,2], Huy Vo[1,3], Cláudio Silva[1,2], and Juliana Freire[1,2]
[1]Center for Urban Science and Progress, New York University
[2]Department of Computer Science and Engineering, New York University
[3]Department of Computer Science, the City College of the City University of New York
{masayo.ota, huy.vo, csilva, juliana.freire}@nyu.edu

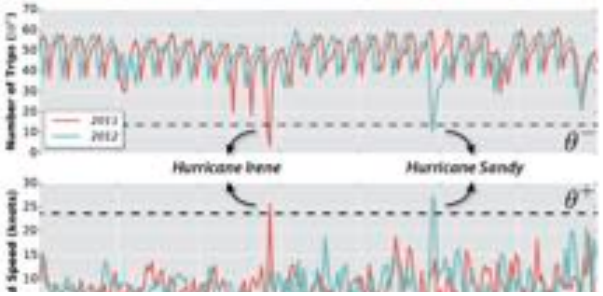
Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets

Fernando Chirigati* Harish Doraiswamy* Theodoros Damoulas*† Juliana Freire*

* New York University * University of Warwick † Alan Turing Institute
{fchirigati,harishd,juliana.freire}@nyu.edu damoulas@warwick.ac.uk

ABSTRACT

The increasing ability to collect data from urban environments, coupled with a push towards openness by governments, has resulted in the availability of numerous spatio-temporal data sets covering diverse aspects of a city. Discovering relationships between these data sets can produce new insights by enabling domain experts to not only test but also generate hypotheses. However, discovering these relationships is difficult. First, a relationship between two data sets may occur only at certain locations and/or time periods. Second, the sheer number and size of the data sets,



Collaborations with social scientists, architects, and city agencies

NYU TANDON SCHOOL OF ENGINEERING