VAST • INFOVIS • SCIVIS
BIOVIS • LDAV

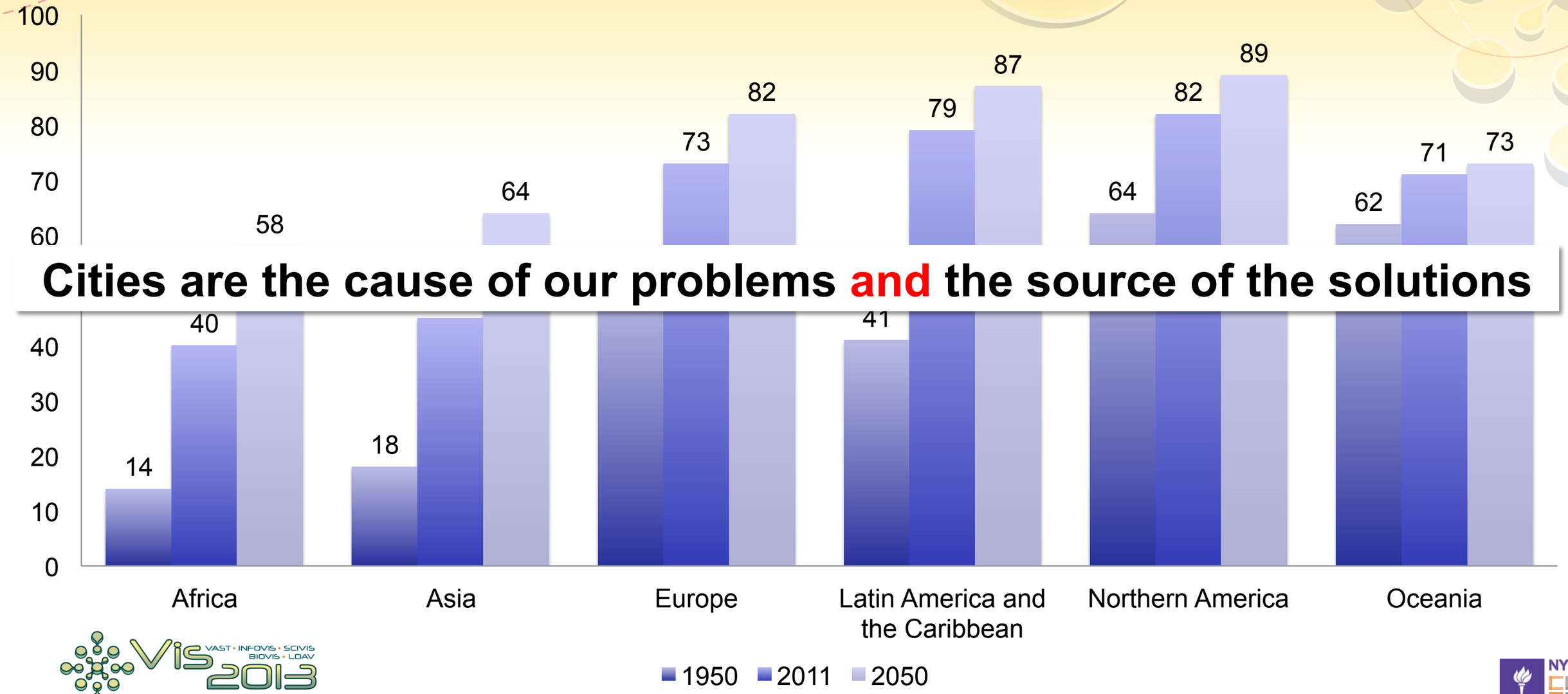
VIS 2013

Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Cab Trips

Nivan Ferreira, Jorge Poco, Huy T. Vo, Juliana Freire, and Claudio T. Silva

Poly and CUSP, New York University

Big Cities – the world is urbanizing



Cities are the cause of our problems and the source of the solutions

VAST · INFOVIS · SCIVIS
BIOVIS · LDAV
VIS 2013

NYU-poly CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

Exploring Urban Data

Infrastructure

Condition, operations

Environment

Meteorology, pollution, noise, flora, fauna

People

Relationships, location, economic /communications activities, health, nutrition, opinions, …

Properly acquired, integrated, and analyzed, data can

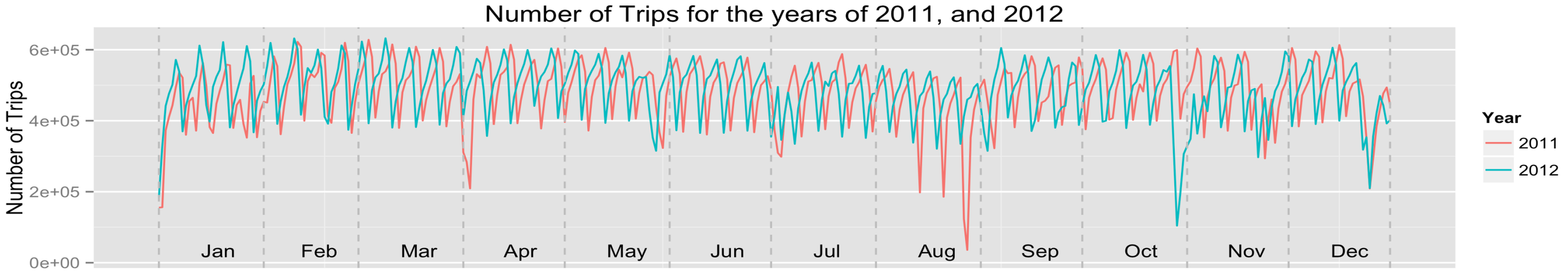
- Take government beyond imperfect understanding
  - Better (and more efficient) operations, better planning, better policy
- Improve governance and citizen engagement
- Enable the private sector to develop new services for citizens, governments, firms
- Enable a revolution in the social sciences

Urban Data Sources

- Organic data flows
  - Administrative records (census, permits, …)
  - Transactions (sales, communications, …)
  - Operational (traffic, transit, utilities, health system, …)
  - New and social media (Twitter feeds, blog posts, Facebook, …)
- Sensors
  - Personal (location, activity, physiological)
  - Fixed *in situ* sensors
  - Crowd sourcing (mobile phones, …)
  - Choke points (people, vehicles)
- Opportunities for “novel” sensor technologies
  - Visible, infrared and spectral imagery
  - RADAR, LIDAR
  - Gravity and magnetic, seismic, acoustic
  - Ionizing radiation, biological, chemical

Taxis as Sensors for NYC

- Taxis are *sensors* that can provide unprecedented insight into city life: economic activity, human behavior, mobility patterns, …
  - *"What is the average trip time from Midtown to the airports during weekdays?''*
  - *"How the taxi fleet activity varies during weekdays?''*
  - *"How was the taxi activity in Midtown affected during a presidential visit?''*
  - *"How did the movement patterns change during Sandy?"*
  - *"Where are the popular night spots?"*



VAST · INFOVIS · SCIVIS
BIOVIS · LDAV
VIS 2013

NYU-poly CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

Exploring Taxi Trips: Challenges

Taxi data are:

- **Big:** 520 million trips -- ~500k trips/day
  - Can't use existing tools for interactive exploration
- **Complex**: Multiple variables: *spatial and temporal + trip attributes*
  - Hard to select data -- too many data slices
- **Dirty:** Taxis in the river…

Domain scientists and decision makers are unable to interactively explore the whole data

Exploring Taxi Trips: Challenges

Taxi data are:

- **Big:** 520 million trips -- ~500k trips/day
  - Can't use existing tools for interactive exploration
- **Complex**: Multiple variables: *spatial and temporal + trip attributes*
  - Hard to select data -- too many data slices
- **Dirty:** Taxis in the river…

7-8am

8-9am

9-10am

10-11am

TaxiVis: Visually Exploring NYC Taxi Data

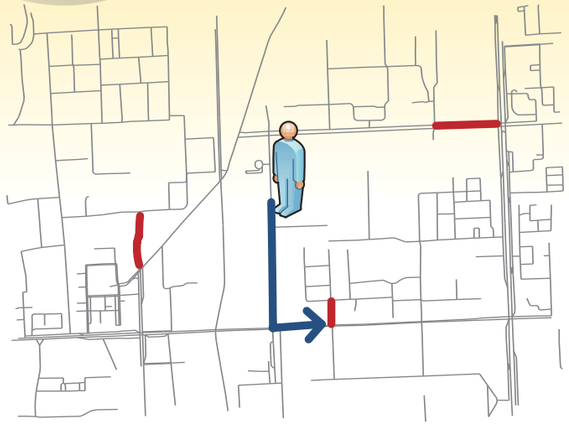
- New model that allows users to visually query taxi trips, easily select and compare different spatial-temporal slices
  - Data selection through visual manipulations
  - Use visualization to explore selected data
- Support for origin-destination queries that enable the study of mobility across the city
- Use multiple coordinated views to allow comparisons, and brushing to support query refinements
- Use of adaptive level-of-detail rendering and heat maps to generate clutter-free visualization for large results
- Scalable system that provides interactive response times for spatio-temporal queries over large data

VAST • INFOVIS • SCIVIS
BIOVIS • LDAV
VIS 2013

NYU-poly
CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

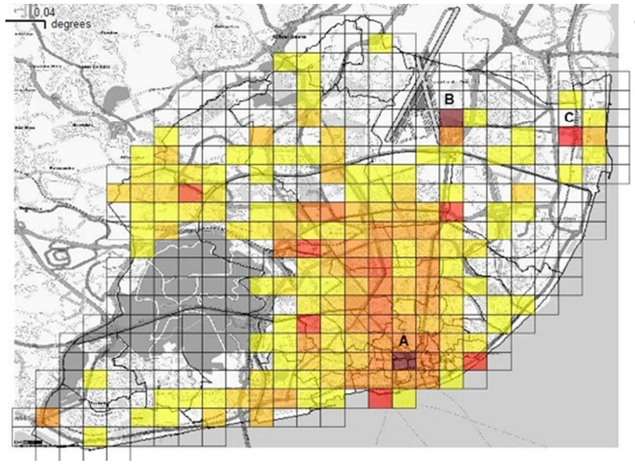
Related Work – taxi data

- Recommendation Systems
  [*Ge et al. 2010*]
  [*Yuan et al. 2011*]



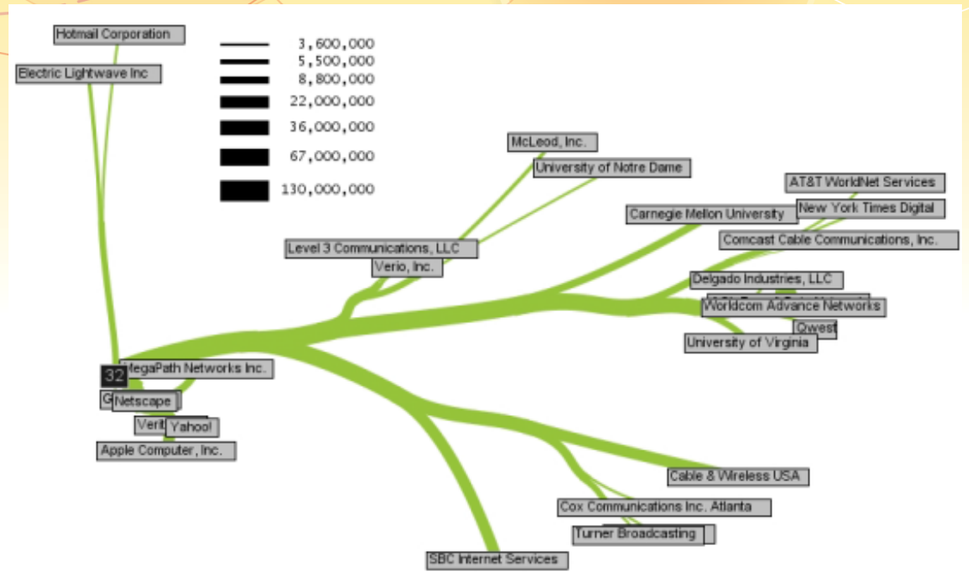
- Land-use Classification
  [*Pan et al. 2013*]

- Human Mobility
  [*Veloso et al. 2011*]
  [*Liang et al. 2012*]
  [*Peng et al. 2012*]



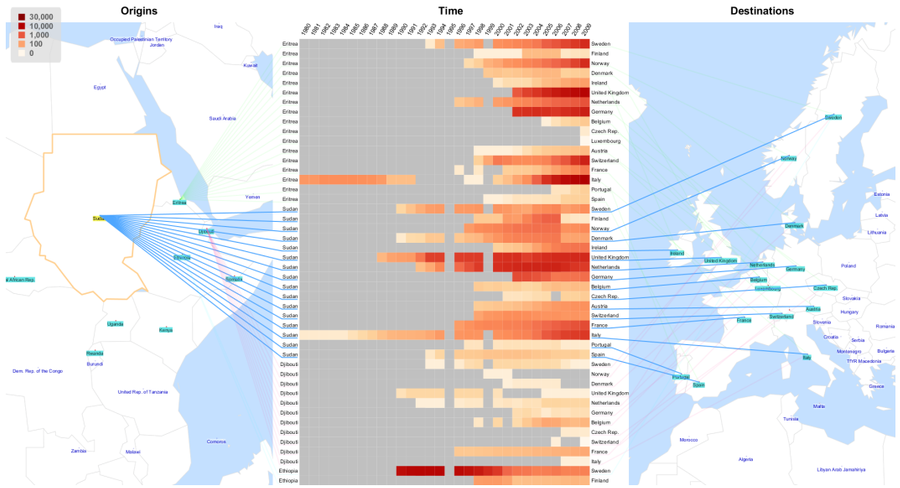
Related Work – mobility visualization

- Flow Maps *[Phan et al. 2005]*



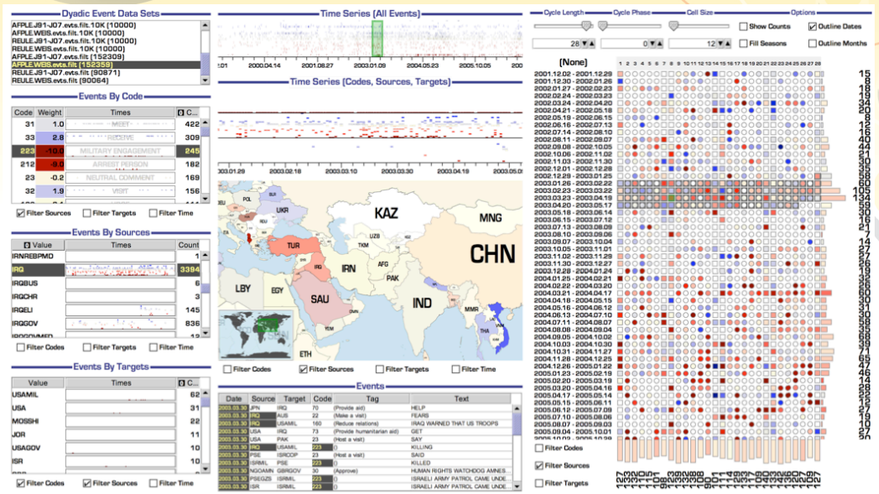
- OD Maps *[Wood et al. 2010]*

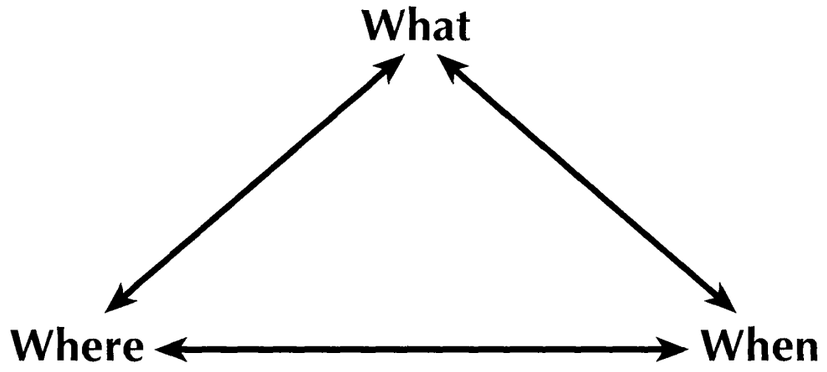
- Flowstrates *[Boyadin el al. 2011]*



Related Work – querying and visualizing spatio-temporal data

- Spatio-temporal querying model *[Peuquet 1994]*
- Cross-filtered views *[Weaver 2008]*
- Visual query languages
  - GeoPQL *[Ferri and Rafanelli 2005]*
  - Moving GeoPQL *[D'Ulizia et al. 2012]*
  - Query-by-trace *[Erwig and Schneider 2000]*
  - TrajectoryLenses *[Krueger et al. 2013]*





VIS 2013 VAST • INFOVIS • SCIVIS BIOVIS • LDAV

NYU-poly CUSP NYU CENTER FOR URBAN SCIENCE+PROGRESS

Desiderata in TaxiVis

- End-to-end solution for interactive visual analytics
  - Couples database back-end with a usable interface front-end
- Support interactive queries for the entire (growing) datasets
  - Out-of-core data access

Data – limits of existing technology

- Raw data:
  - 3 years: 2009, 2011, and 2012
  - 150 GB in 48 CSV files
  - 520M trips total
- After cleanup and transformation:
  - 50GB in binary format
  - 12 fields with 2 temporal spatial attributes

	SQLite	PostgreSQL
Storage Space in GB	100	200
Building Indices in Minutes (One Year of Data)	3,120	780
1K Items Query in Seconds	8	3
100K Items Query in Seconds	85	24

Data – limits of existing technology

- Raw data:
  - 3 years: 2009, 2011, and 2012
  - 150 GB in 48 CSV files
  - 520M trips total
- After cleanup and transformation:
  - 50GB in binary format
  - 12 fields with 2 temporal spatial attributes
- Solution:
  - New spatio-temporal index based on out-of-core Kd-tree
    - Can also index other attributes

	SQLite	PostgreSQL	Our Solution
Storage Space in GB	100	200	30
Building Indices in Minutes (One Year of Data)	3,120	780	28
1K Items Query in Seconds	8	3	0.2
100K Items Query in Seconds	85	24	2

Data Exploration: A Two-Phase Process

Data selection

- Specify query constraints

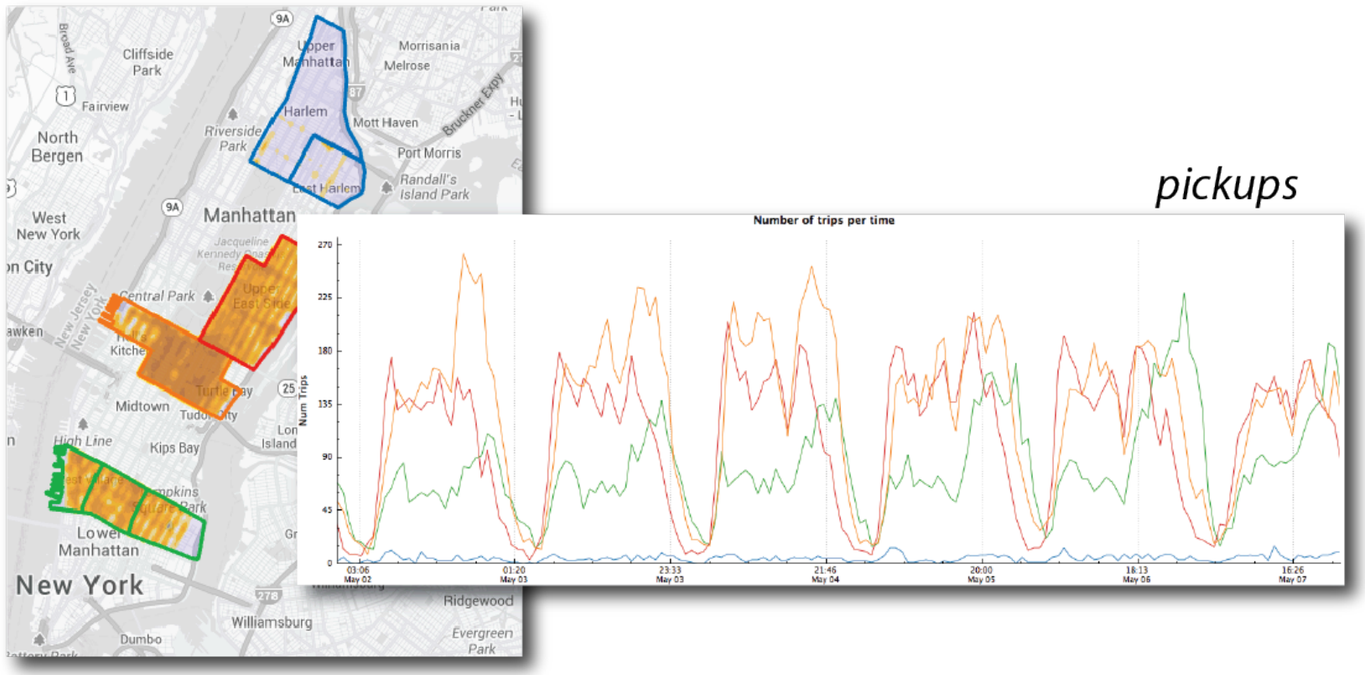
Visual analysis

- Investigate selected data through visualization
- Discover regions of interest
- Define new data selections for further exploration

We unify the two phases of the process through visual operations

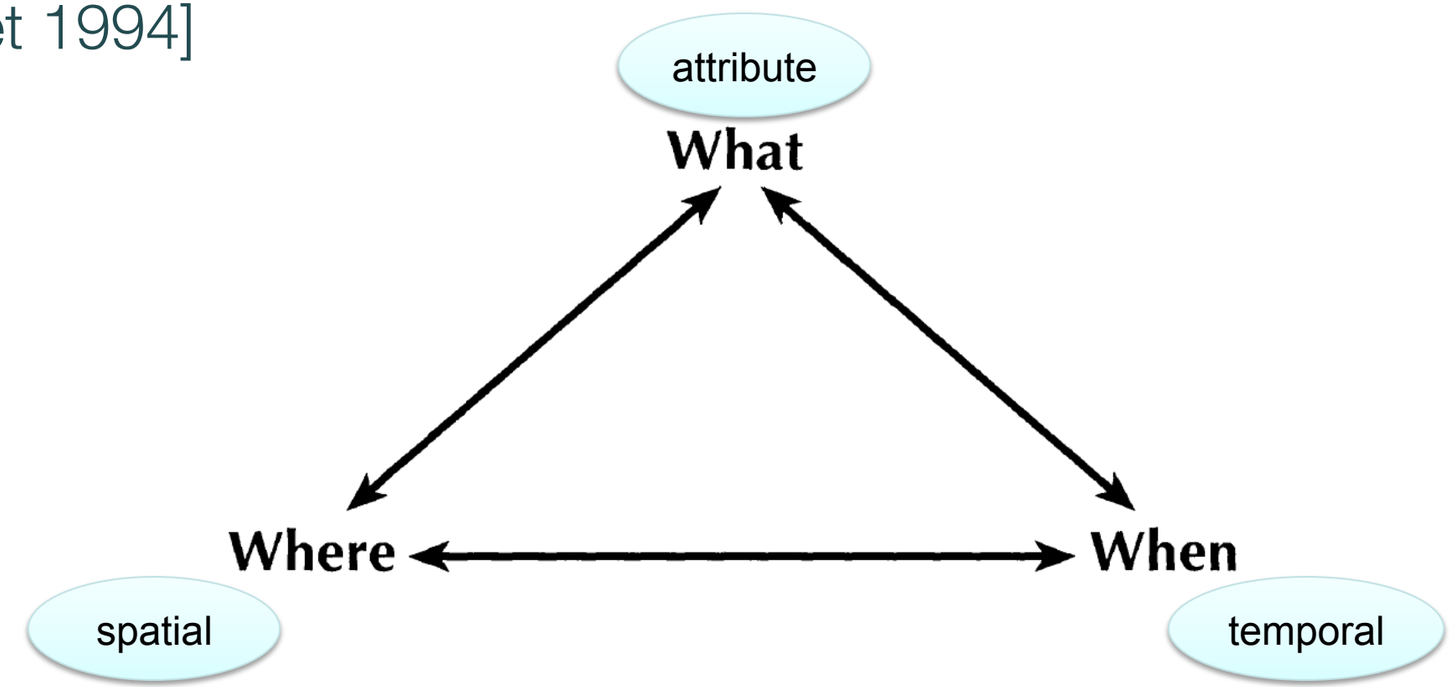
Interactive Visual Exploration

- Help users easily select data slices
  - Composition of spatio-temporal constraints
- Provide visualizations of the slices within the spatio-temporal context
- Multiple coordinated views
  - Time series, histogram plots
  - Heat maps
- Comparative visualizations
  - Multiple query results
  - Exploration in time
  - Summary of attributes



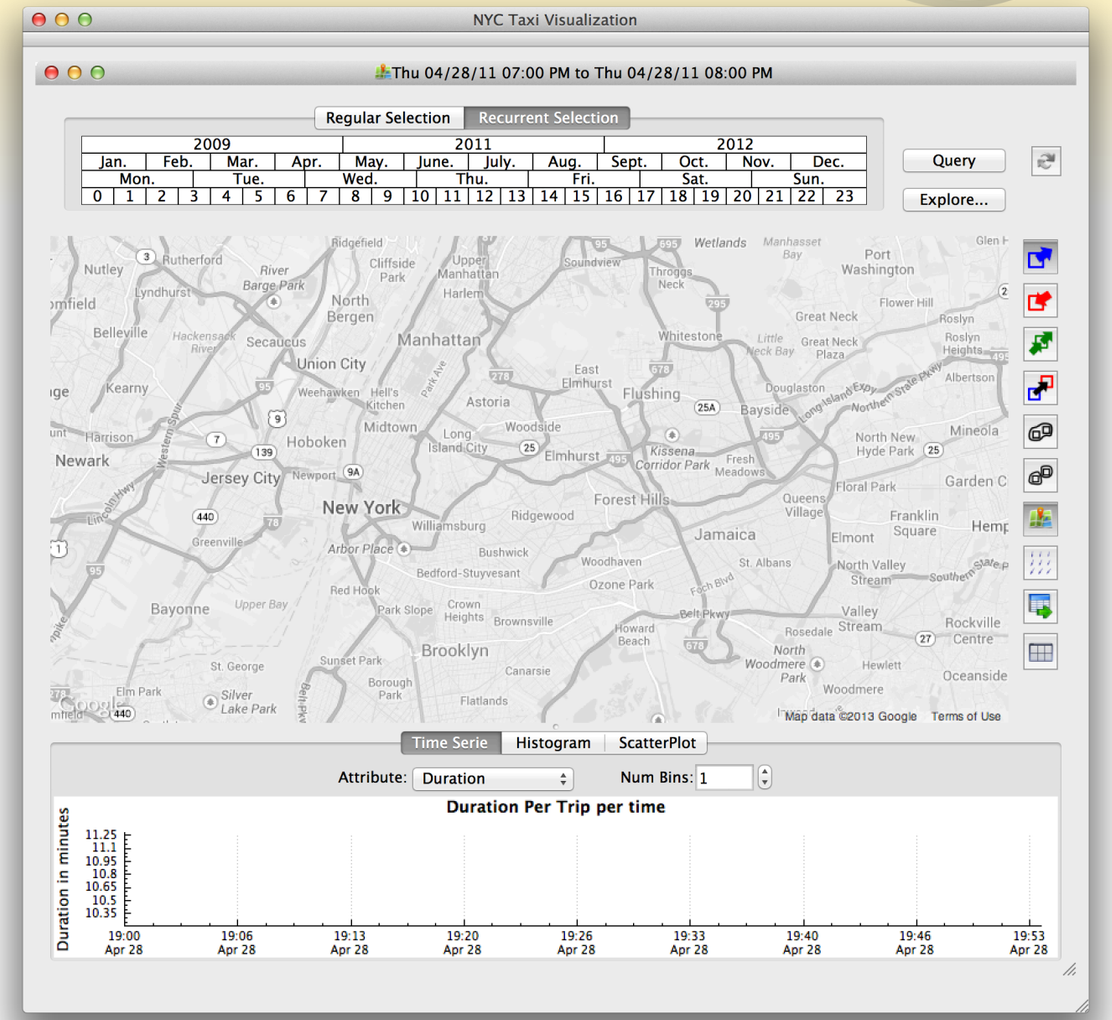
Visual Query Model

- Data selection by visual operations
- Each data slice can be assigned a different visual representation
  - Spatial context is maintained in the map view
- Query Expressiveness [Peuquet 1994]
  - when + where ➔ what
  - when + what ➔ where
  - where + what ➔ when



When + Where → What

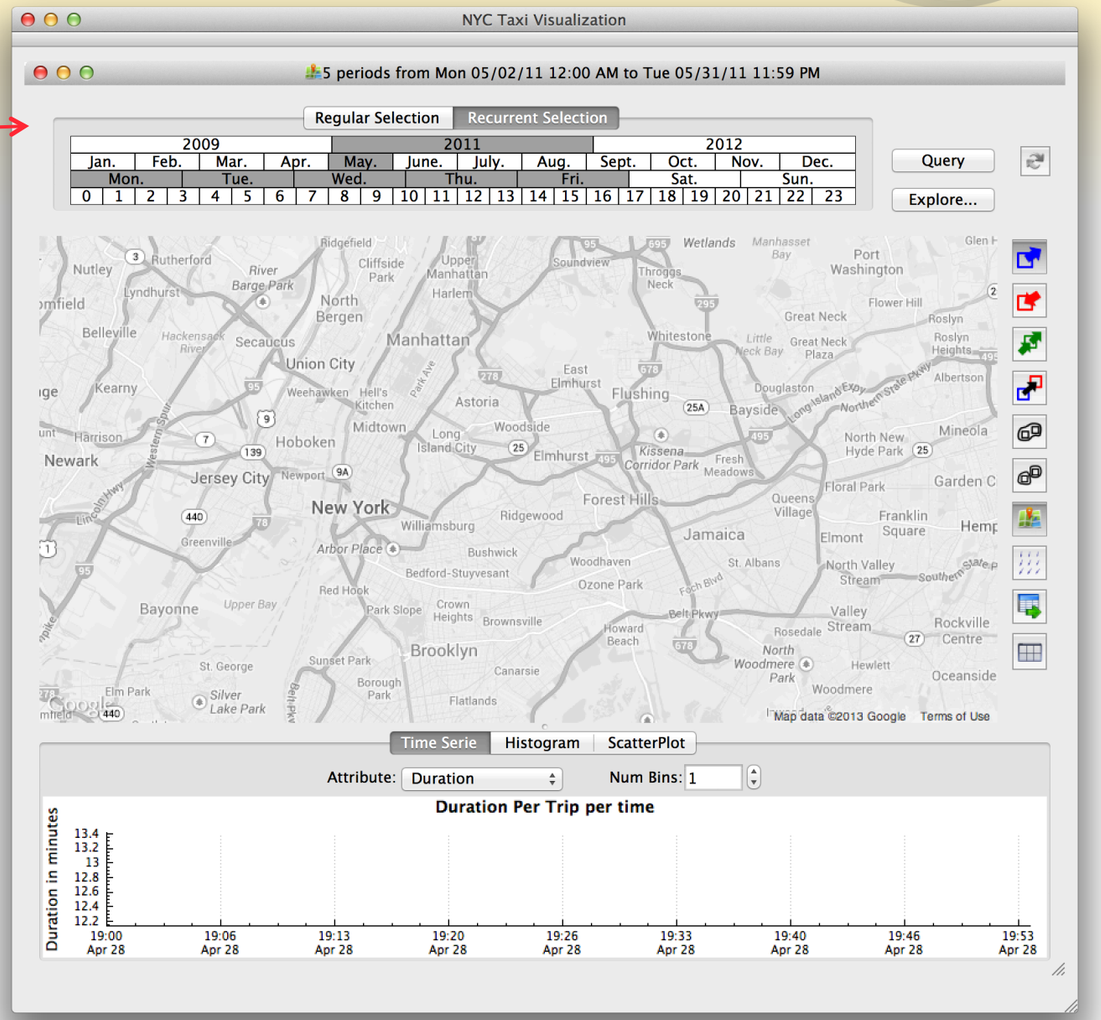
"What is the average trip time from Midtown to the airports during weekdays?



When + Where → What

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

When?

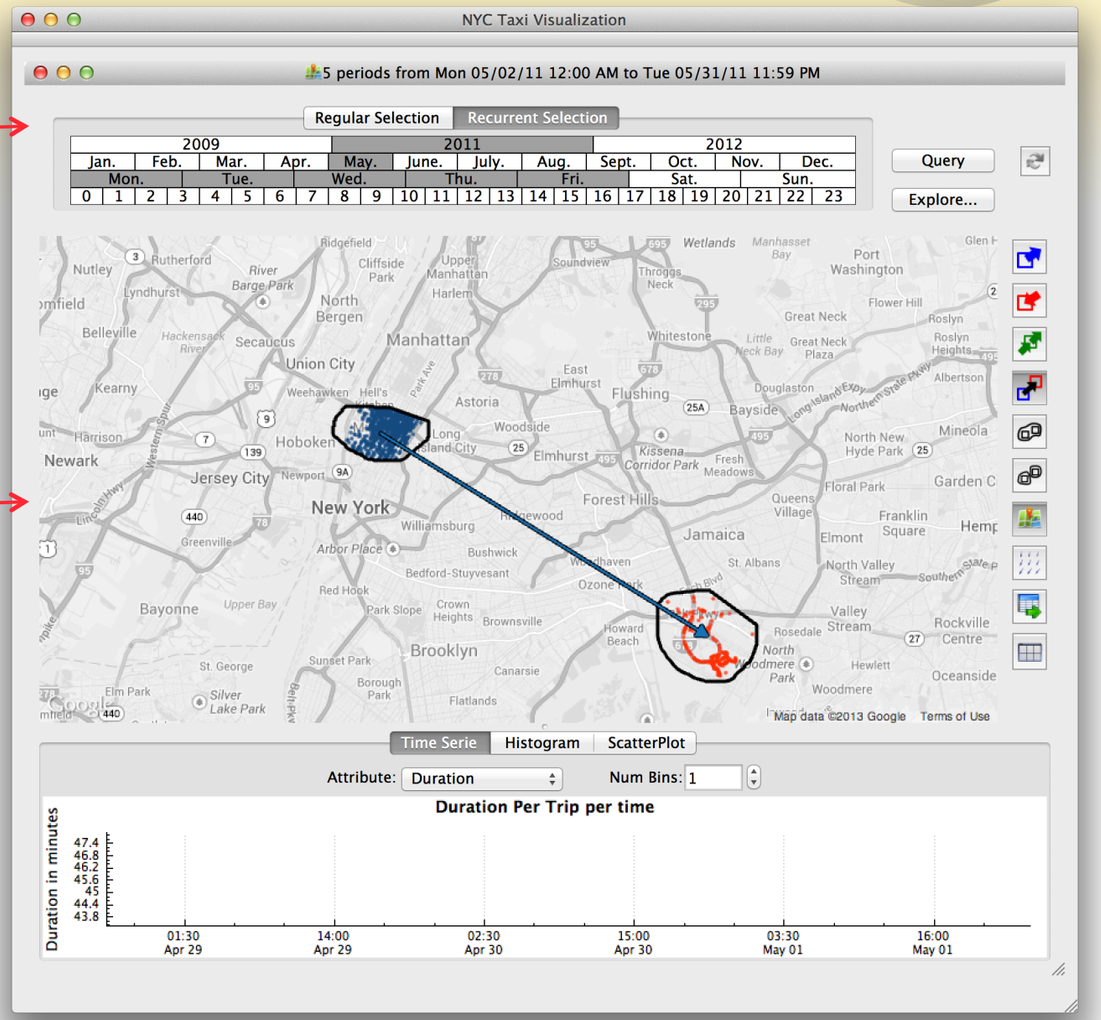


When + Where → What

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

When?

Where?



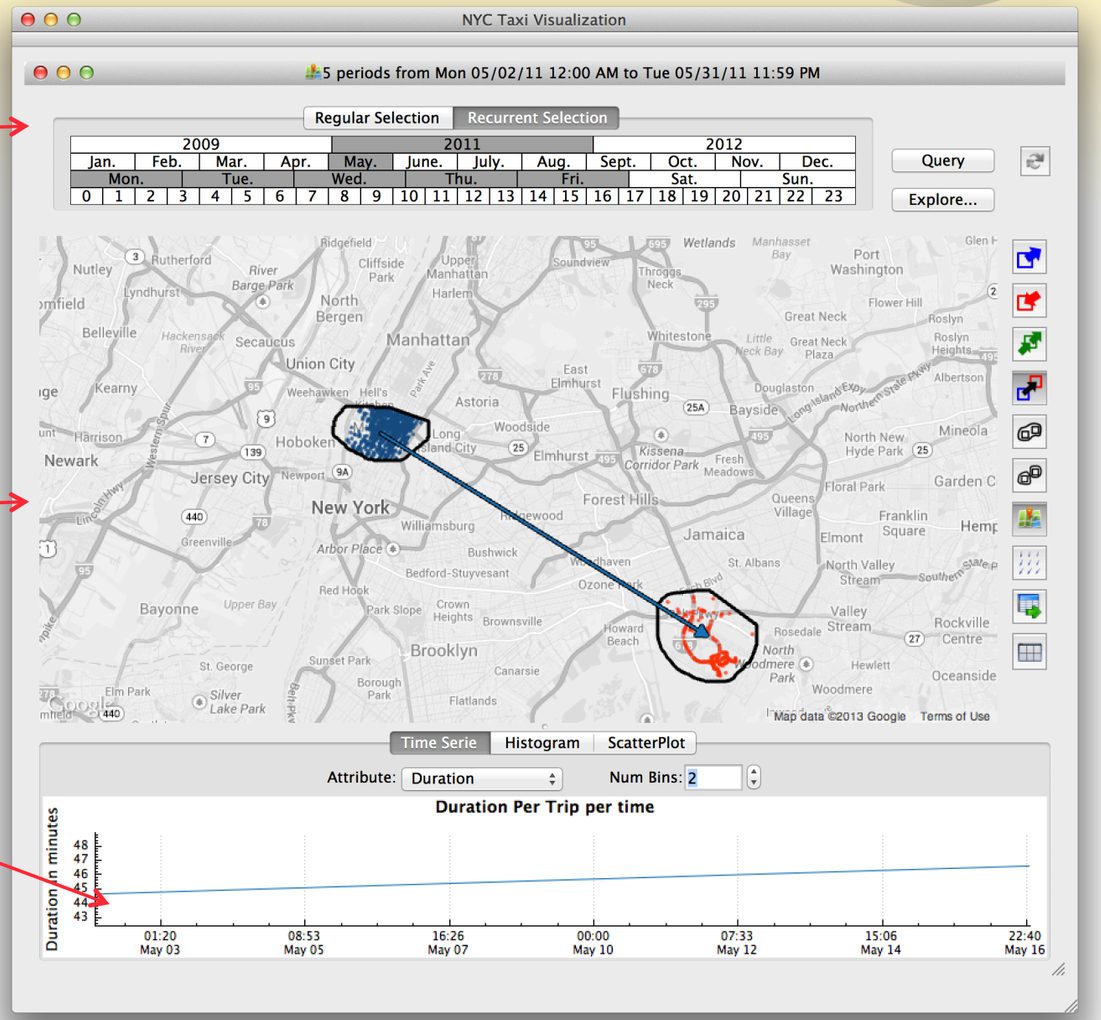
When + Where → What

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

When?

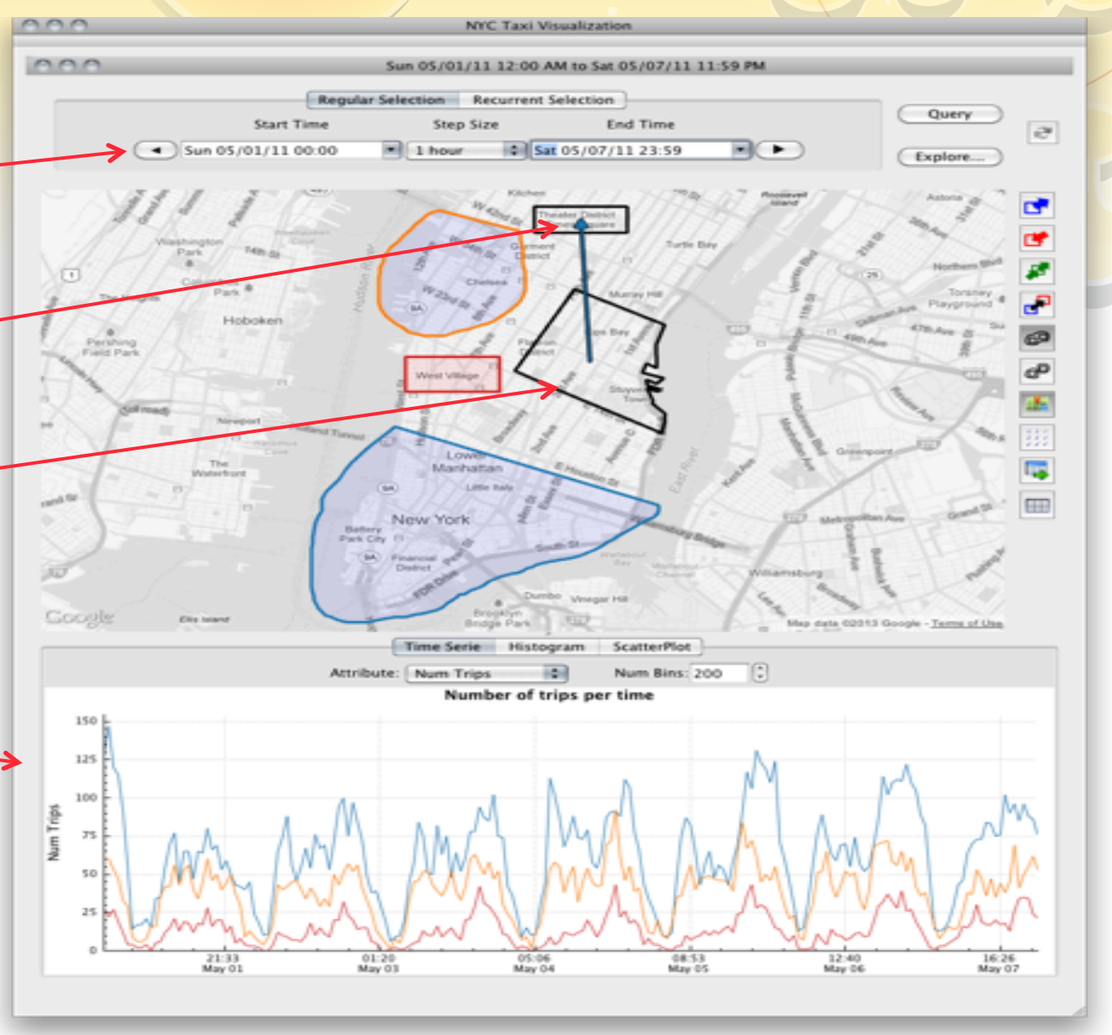
Where?

What



Visual Data Selection

```
SELECT  *
FROM    trips
WHERE pickup_time in (5/1/11,5/7/11)
          AND
      dropoff_loc in “Times Square”
          AND
      pickup_loc    in “Gramercy”
```

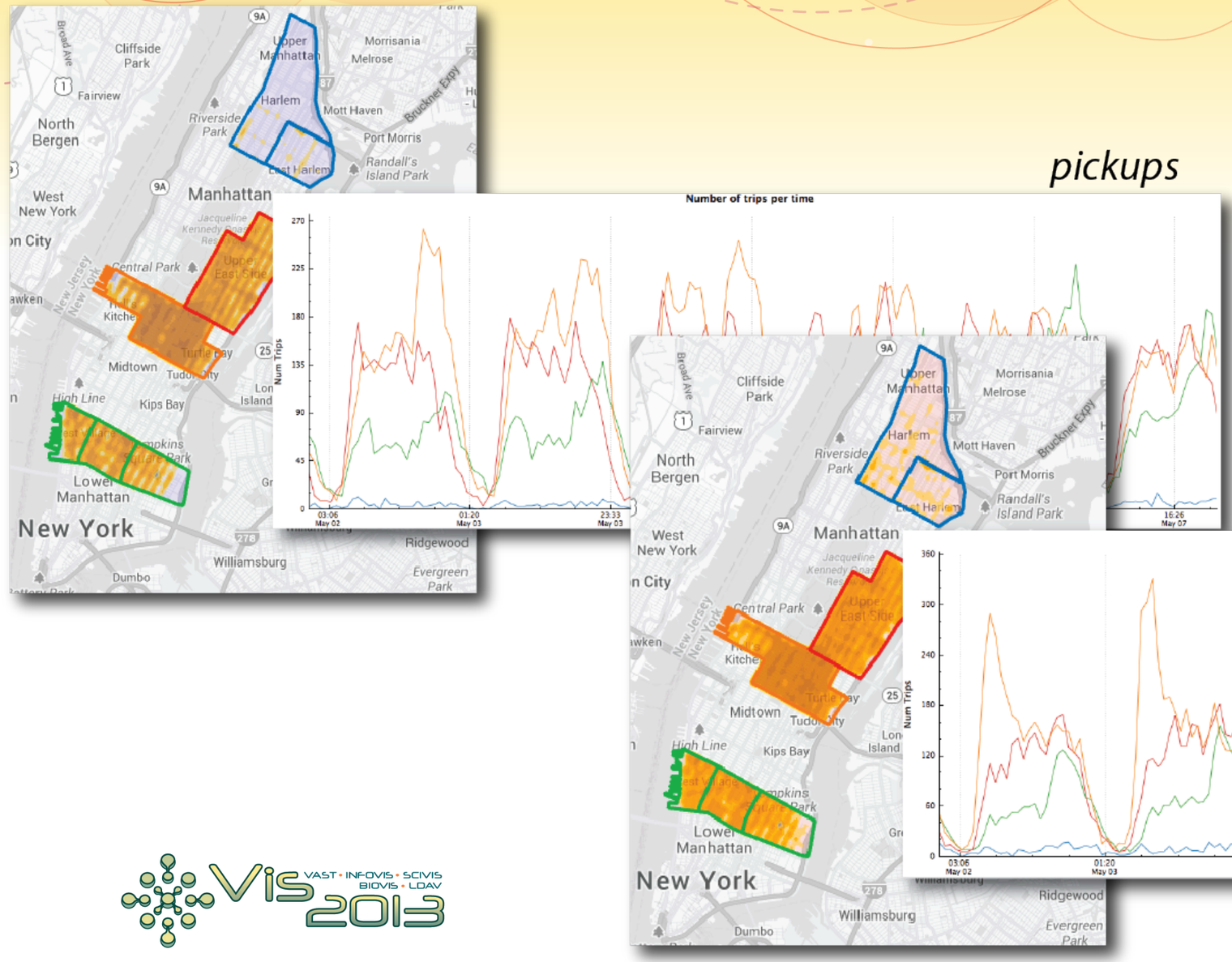


Interactively explore data through the map view and plot widgets

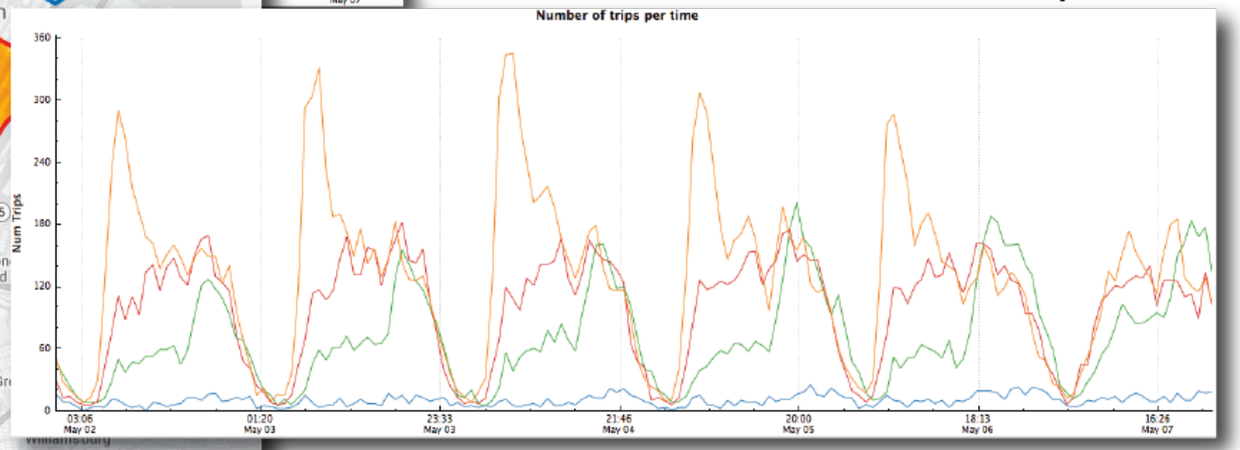
TaxiVis in Action

Cross-filtered Views: trip distributions by area

pickups



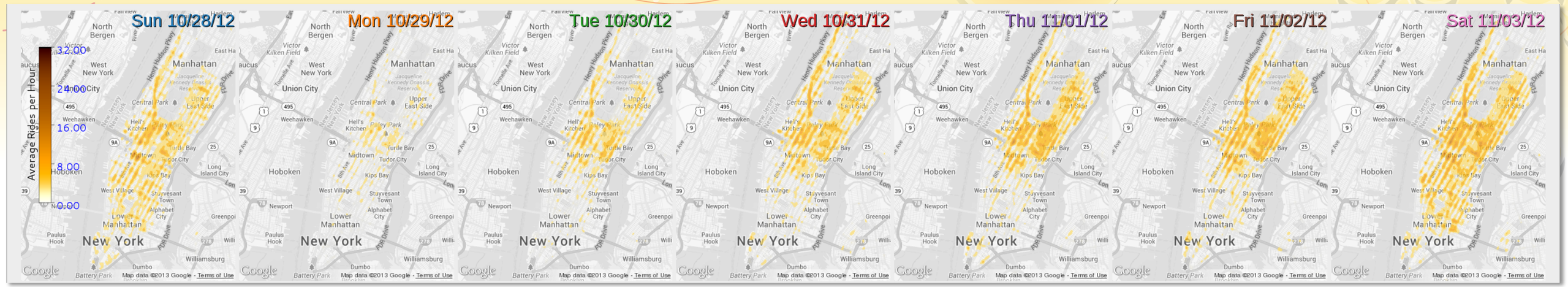
dropoffs

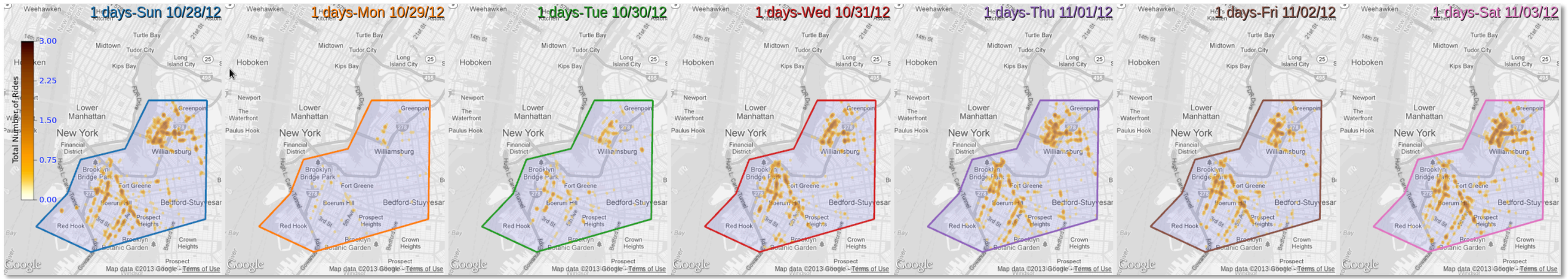


VIS 2013 VAST • INFOVIS • SCIVIS BIOVIS • LDAV

NYU-poly CUSP NYU CENTER FOR URBAN SCIENCE•PROGRESS

Temporal Comparison: Hurricane Sandy

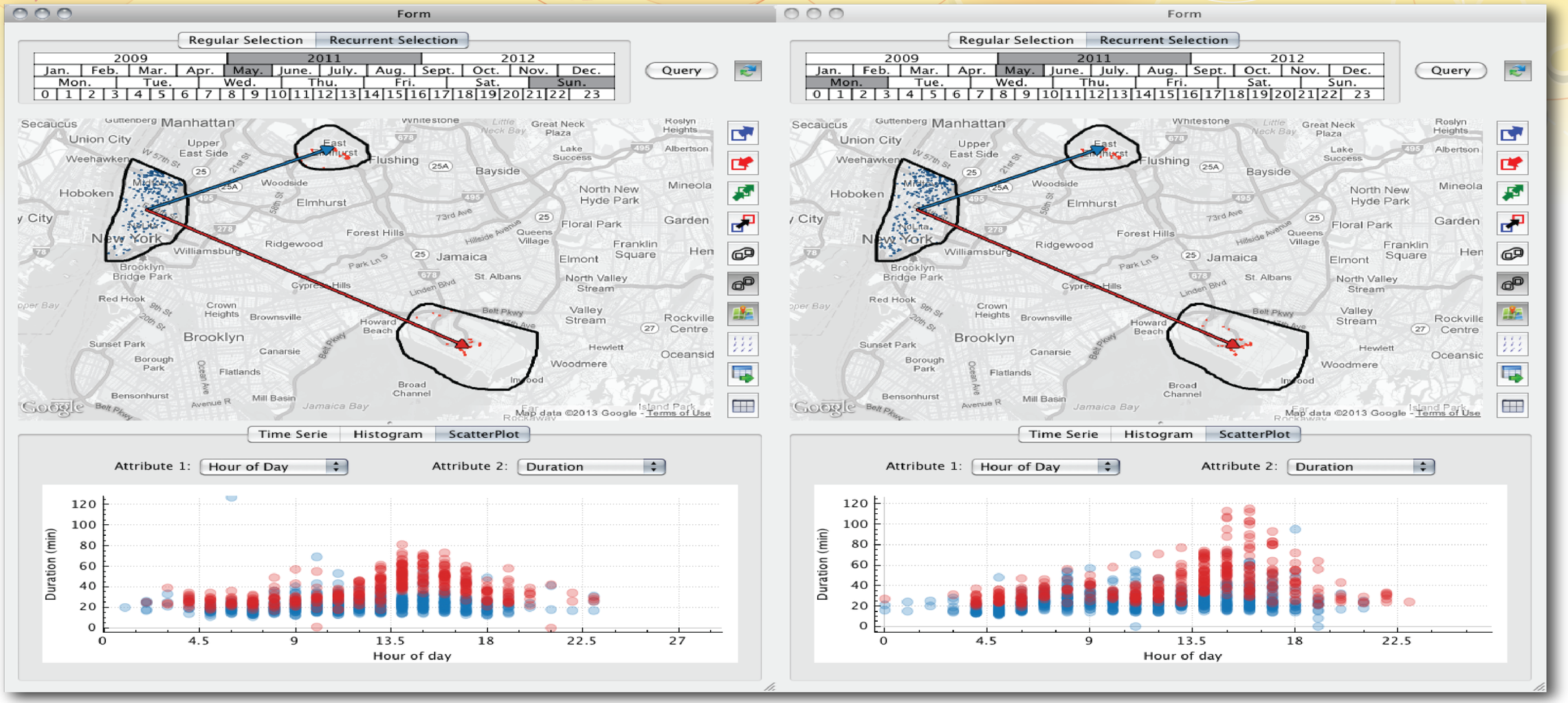




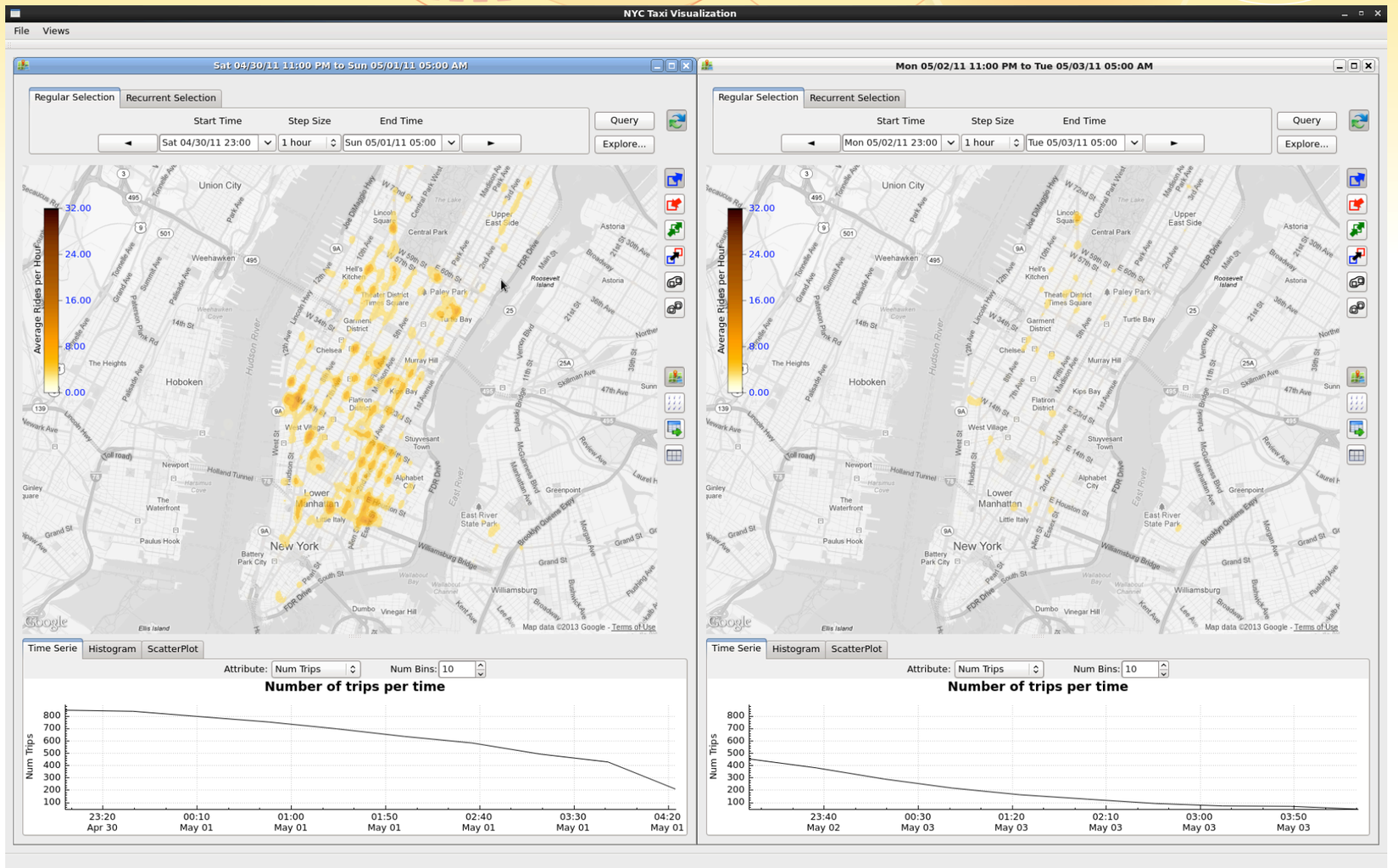
VAST • INFOVIS • SCIVIS
BIOVIS • LDAV
VIS 2013

NYU-poly
CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

Analyzing Movement



Comparative Visualization: Night Life in NYC



Saturday

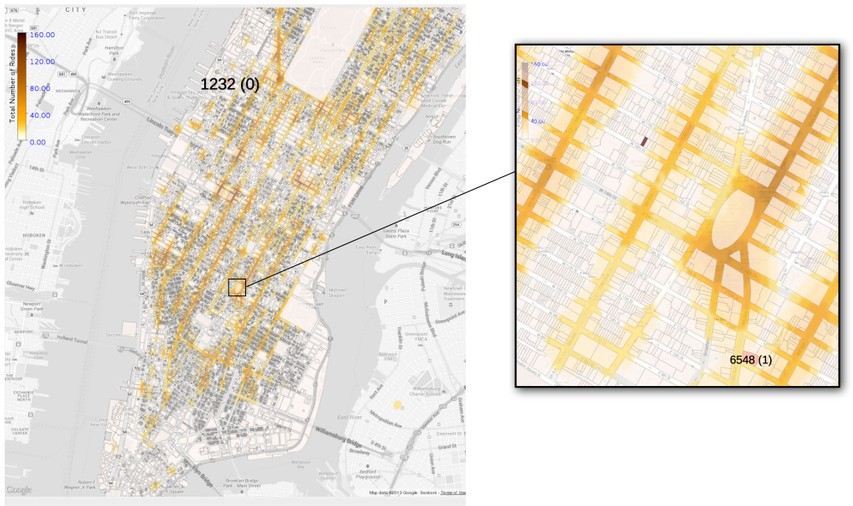
Monday

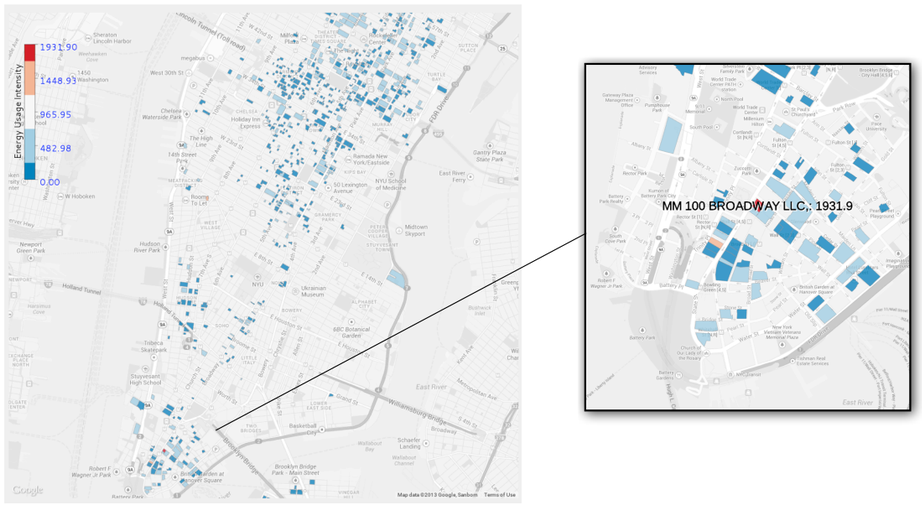
Conclusion & Limitations

- Easy-to-use system to interactively explore large multivariate spatial-temporal data
  - Couples database and visualization
- Desktop-based application
  - Web-based map
- Need support for multiple data layers

Future and Ongoing Work

- Apply to other urban mobility data, e.g., data from the NYC bike share program
- Support additional data layers: weather, gas prices, news, tweets, etc.
- Utilize parallel processing





VAST • INFOVIS • SCIVIS
BIOVIS • LDAV
VIS 2013

NYU-poly
CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

Acknowledgements

- The Taxi & Limousine Commission of New York City
- National Science Foundation (CNS-1229185, CNS-1153503, IIS 1139832, IIS-0905385, IIS-1142013, AGS 0835821)
- The Department of Energy

VAST • INFOVIS • SCIVIS
BIOVIS • LDAV
VIS 2013

NYU-poly
CUSP
NYU CENTER FOR URBAN SCIENCE+PROGRESS

Thanks!


VAST • INFOVIS • SCIVIS
BIOVIS • LDAV
VIS 2013


NYU-poly
CUSP
NYU
CENTER FOR URBAN
SCIENCE•PROGRESS