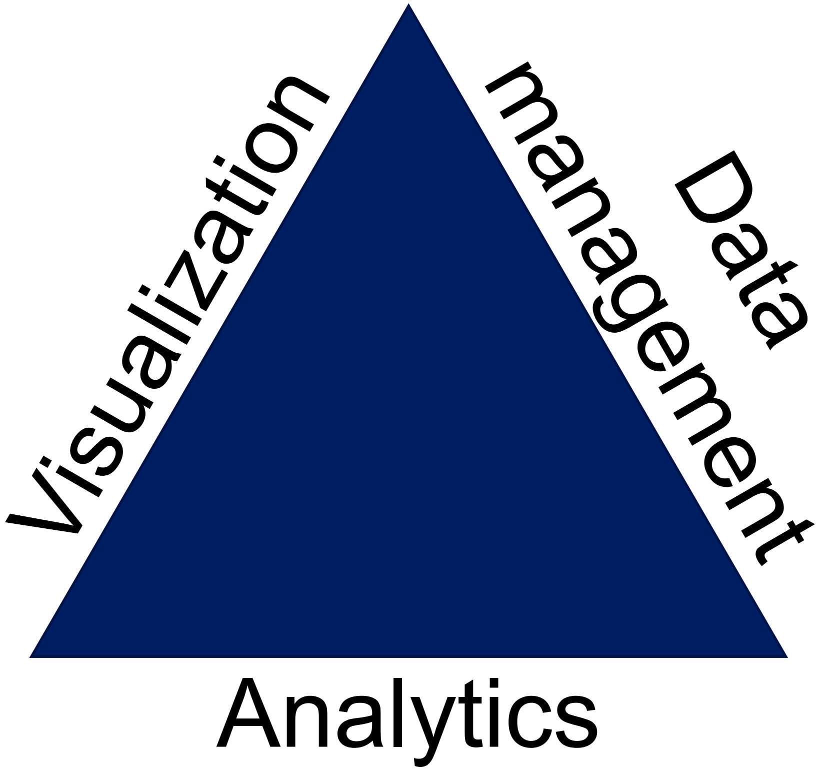
Advanced spatial data visualization

CS424: Visualization & Visual Analytics

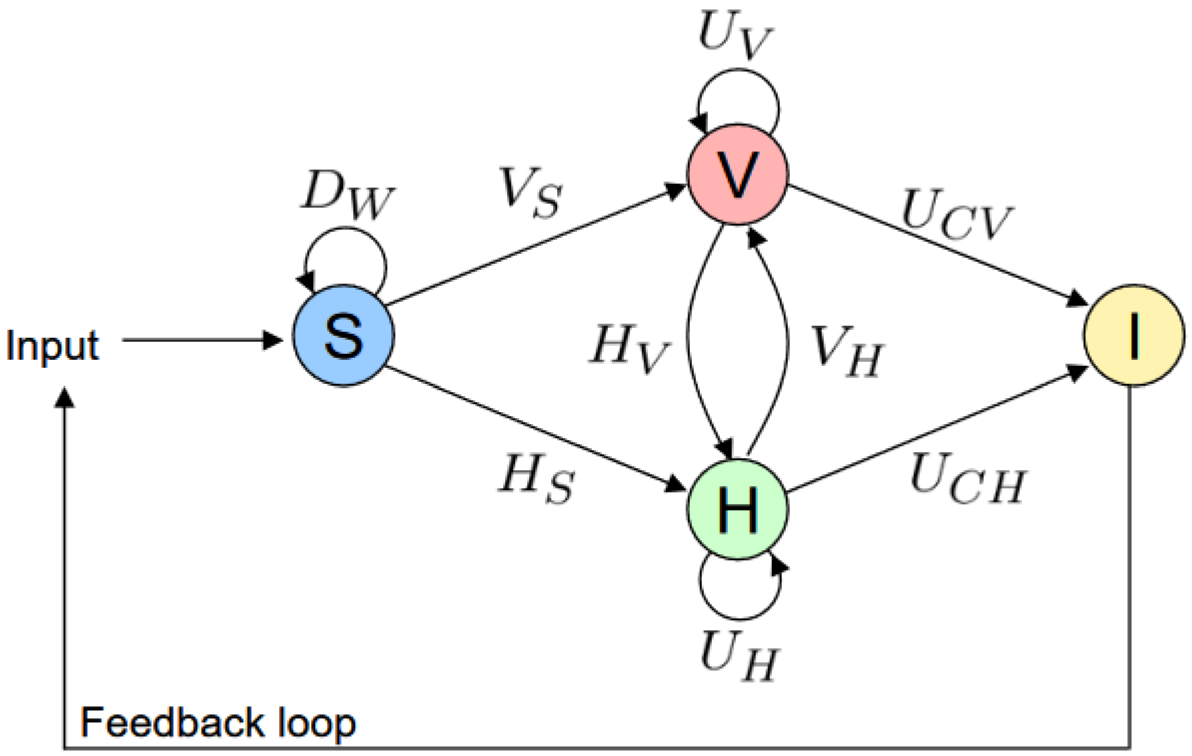
Fabio Miranda

https://fmiranda.me

Visual analytics



Visual analytics process



["Visual Analytics: Scope and Challenges", Kleim et al., 2008]

$F: S \rightarrow I$
$f \in \{D_w, V_x, H_y, U_z\}$
S: data, I: insight

D_w: data pre-processing functionalities (data transformation, data cleaning, data selection, data integration)

$V_x, x \in \{S, H\}$: visualization functions
- V_S: functions visualizing data ($S \rightarrow V$)
- V_H: functions visualizing hypotheses ($\mathrm{H} \rightarrow V$)

$H_y, y \in \{S, V\}$: hypotheses generation process
- H_S: hypotheses from automatic process ($S \rightarrow H$)
- H_v: hypotheses from visualization ($\mathrm{V} \rightarrow H$)

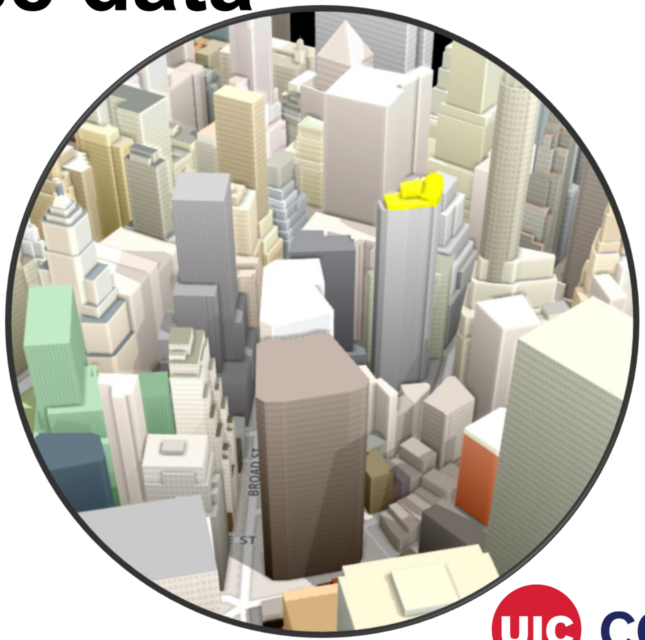
$U_z, z \in \{V, H, CV, CH\}$: user interactions
- U_V: visualization interaction ($V \rightarrow V$)
- U_H: hypotheses interaction ($\mathrm{H} \rightarrow H$)

U_{cv}: insight concluded from visualization
U_{ch}: insight concluded from hypothesis

Diverse data

Diverse data

Diverse data



Diverse data

Occupational therapists

Urban noise experts

Transportation engineers

Public health experts

Environmental scientists

Architects

Urban planners

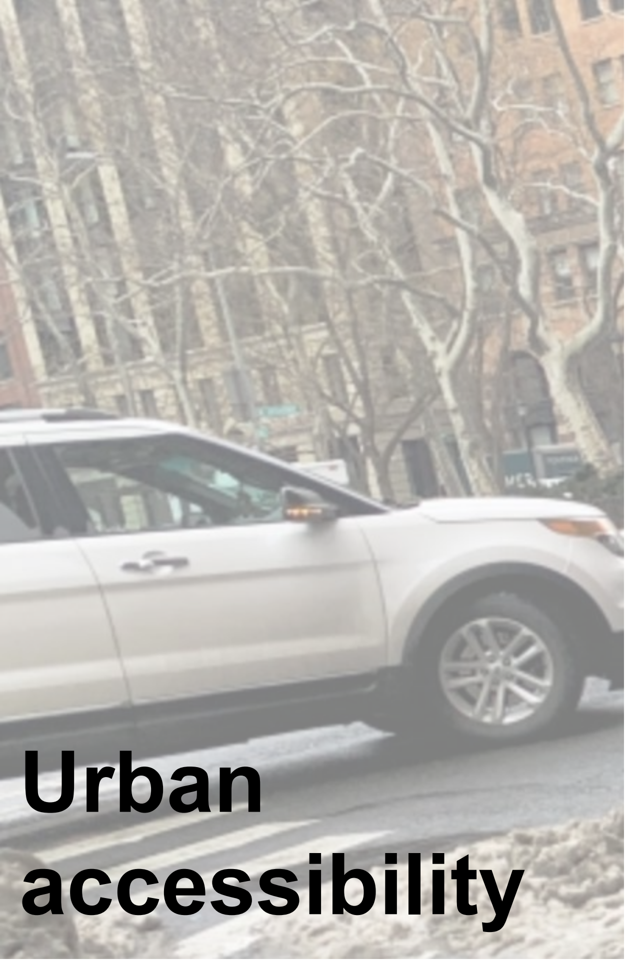
Diverse users

Diverse data

Diverse users

Diverse data

Diverse problems

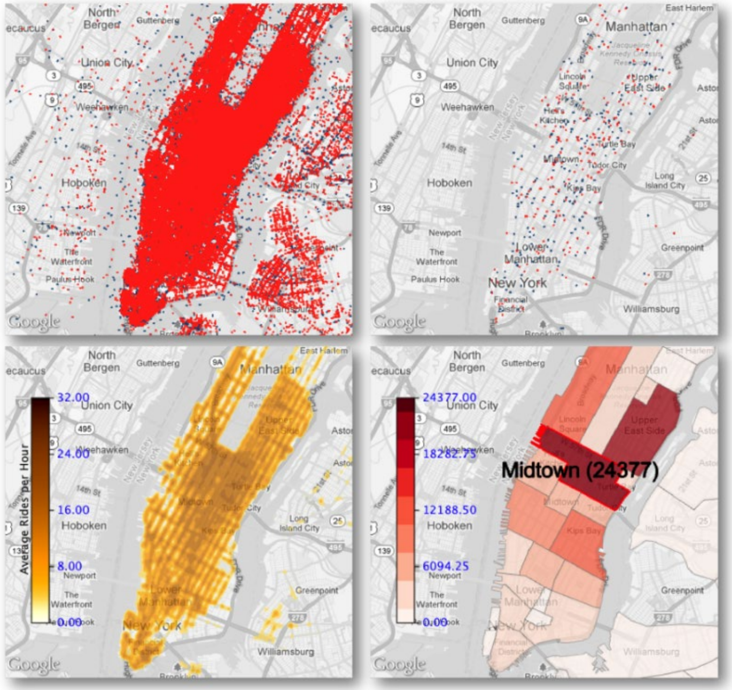


Diverse problems

Diverse users

Diverse data

Visual analysis process: taxi data



Distribution of NYC Taxi Pickups and Dropoffs in Manhattan

Visual analysis process: taxi data

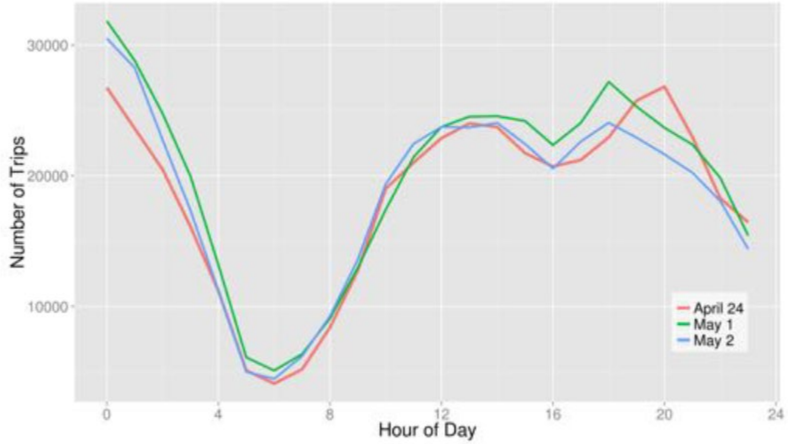
- Data: ~500,000 trips / day; ~1,000,000,000 trips in 5 years
  - Spatiotemporal: pickup + dropoff
  - Trip attributes: distance traveled, fare, tip, etc.
- Government, policy makers and scientists are usually unable to *interactive* explore the *whole* data.
  - Too many slices to examine.

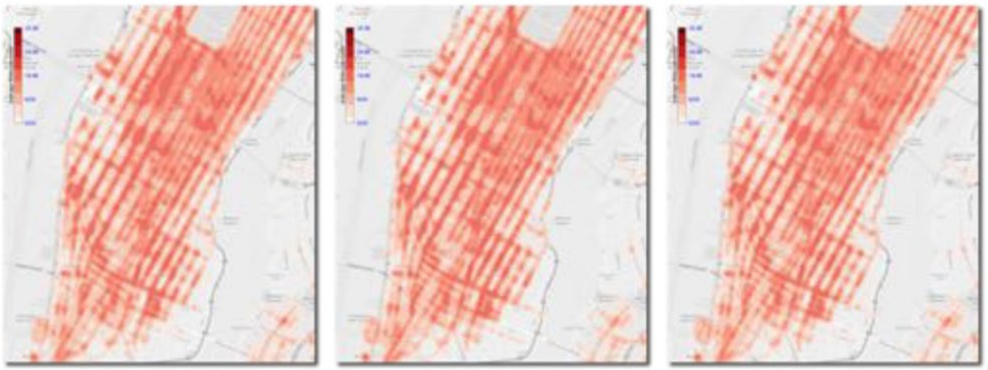
Visual analysis process: taxi data

- Taxi pickups and drop-offs: data sets $S_1, S_2, \dots, S_n$
- Pre-processing
  - Data cleaning (e.g., remove points outside NYC): $d_c(S_1, \dots, S_n)$
  - Data integration (e.g., add weather column): $d_i(S_1, \dots, S_n)$
  - Data transformation (e.g., raw csv's to binary): $d_t(S_1, \dots, S_n)$
  - Data selection (e.g., only Manhattan trips): $d_s(S_1, \dots, S_n)$

Visual analysis process: taxi data

- Too many slices
- 365*24 1-hour slices in just one year
- Which slices are interesting?

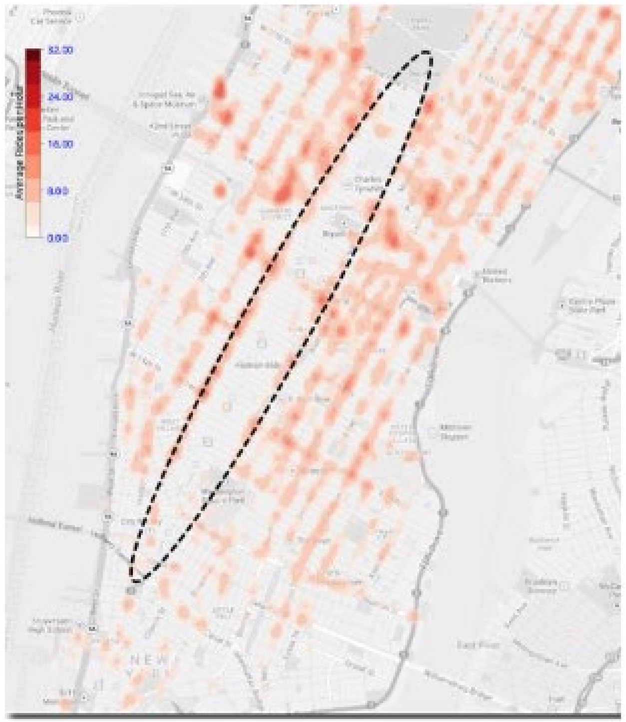




April 24 May 1 May 8

Visual analysis process: taxi data

- Too many slices
- 365*24 1-hour slices in just one year
- Which slices are interesting?



May 1 (8-9am)

Visual analysis process: taxi data

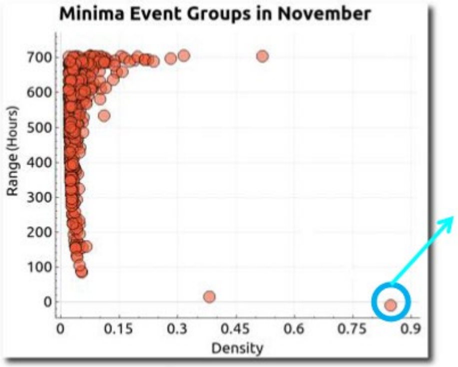
- Too many slices
- 365*24 1-hour slices in just one year
- Which slices are interesting?

- Guide users towards *interesting* data slices
- Data mining:
  - Identifying events with arbitrary spatial structure at multiple temporal scales:

$$h' = h_s(S')$$
$$with\ S' = d_s(d_t(d_i(d_c(S_1, \dots, S_n))))$$

Visual analysis process: taxi data

- Visualization:
  - Visualizing the previous hypothesis: $v' = v_h(h')$
  - Adjusting parameters: $v'' = u_v(v')$
  - Hypotheses from visualization: $h'' = h_v(v'')$, leading to an insight (unusual taxi activity around Macy's parade).



Macy's parade

Visual analysis process: taxi data

- Taxi pickups and drop-offs: data sets $S_1, S_2, \dots, S_n$
- Pre-processing
  - Data cleaning (e.g., removing): $d_c(S_1, \dots, S_n)$
  - Data integration (e.g., adding extra attributes): $d_i(S_1, \dots, S_n)$
  - Data transformation (e.g., format suitable for analysis): $d_t(S_1, \dots, S_n)$
  - Data selection (e.g., subset of sensors): $d_s(S_1, \dots, S_n)$
- Data mining:
  - Identifying events with arbitrary spatial structure at multiple temporal scales:

$$h' = h_s(S'), with\ S' = d_s\left(d_t\left(d_i\left(d_c(S_1, \dots, S_n)\right)\right)\right)$$

- Visualization:
  - Visualizing the previous hypothesis: $v' = v_h(h')$
  - Adjusting parameters: $v'' = u_v(v')$
  - Hypotheses from visualization: $h'' = h_v(v'')$

$$h_v(u_v(v_h(h_s(d_s(d_t(d_i(d_c(S_1, \dots, S_n))))))))$$

Visual analytics mantra

- Visualization mantra:

"*Overview first, zoom / filter, details on demand*"

[Shneiderman, 1996]

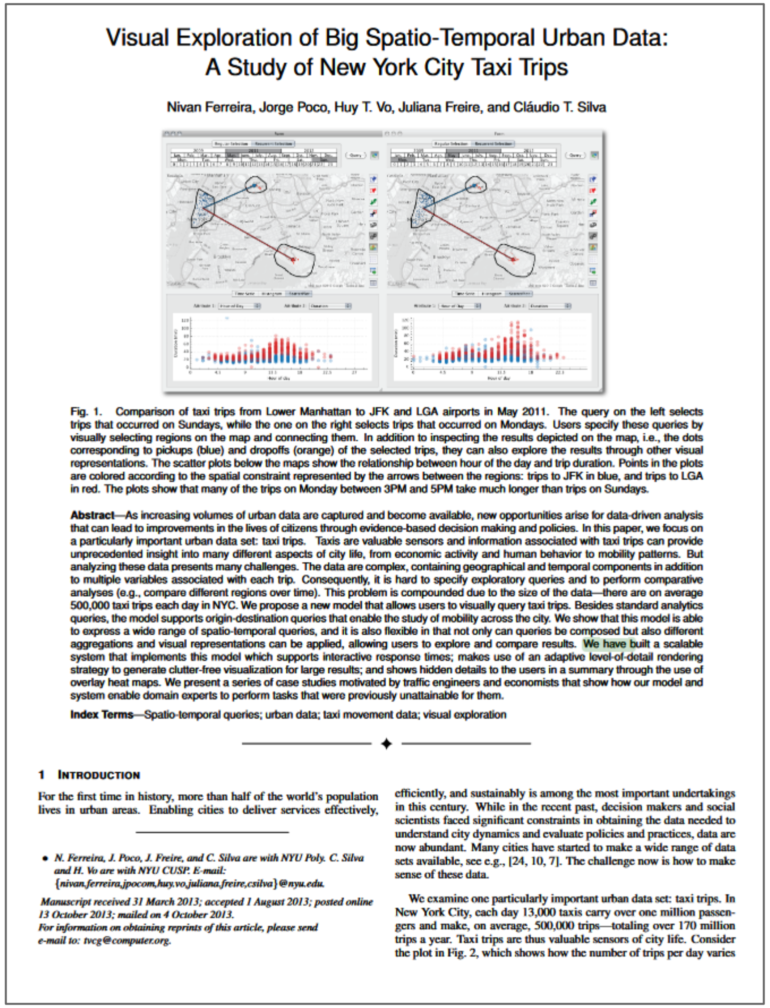
- Visual analytics mantra:

"*Analyze first - show the important - zoom, filter and analyze further - details on demand*"

[Kleim et al., 2008]

Visual analytics + big data

“We propose a new model that allows users to visually query taxi trips. Besides standard analytics queries, the model supports origin-destination queries that enable the study of mobility across the city. We show that this model is able to express a wide range of spatiotemporal queries, and it is also flexible in that not only can queries be composed but also different aggregations and visual representations can be applied, allowing users to explore and compare results.”



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

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

Fig. 1. Comparison of taxi trips from Lower Manhattan to JFK and LGA airports in May 2011. The query on the left selects trips that occurred on Sundays, while the one on the right selects trips that occurred on Mondays. Users specify these queries by visually selecting regions on the map and connecting them. In addition to inspecting the results depicted on the map, i.e., the dots corresponding to pickups (blue) and dropoffs (orange) of the selected trips, they can also explore the results through other visual representations. The scatter plots below the maps show the relationship between hour of the day and trip duration. Points in the plots are colored according to the spatial constraint represented by the arrows between the regions: trips to JFK in blue, and trips to LGA in red. The plots show that many of the trips on Monday between 3PM and 5PM take much longer than trips on Sundays.

Abstract—As increasing volumes of urban data are captured and become available, new opportunities arise for data-driven analysis that can lead to improvements in the lives of citizens through evidence-based decision making and policies. In this paper, we focus on a particularly important urban data set: taxi trips. Taxis are valuable sensors and information associated with taxi trips can provide unprecedented insight into many different aspects of city life, from economic activity and human behavior to mobility patterns. But analyzing these data presents many challenges. The data are complex, containing geographical and temporal components in addition to multiple variables associated with each trip. Consequently, it is hard to specify exploratory queries and to perform comparative analyses (e.g., compare different regions over time). This problem is compounded due to the size of the data—there are on average 500,000 taxi trips each day in NYC. We propose a new model that allows users to visually query taxi trips. Besides standard analytics queries, the model supports origin-destination queries that enable the study of mobility across the city. We show that this model is able to express a wide range of spatio-temporal queries, and it is also flexible in that not only can queries be composed but also different aggregations and visual representations can be applied, allowing users to explore and compare results. We have built a scalable system that implements this model which supports interactive response times; makes use of an adaptive level-of-detail rendering strategy to generate clutter-free visualization for large results; and shows hidden details to the users in a summary through the use of overlay heat maps. We present a series of case studies motivated by traffic engineers and economists that show how our model and system enable domain experts to perform tasks that were previously unattainable for them.

Index Terms—Spatio-temporal queries; urban data; taxi movement data; visual exploration

1 INTRODUCTION

For the first time in history, more than half of the world's population lives in urban areas. Enabling cities to deliver services effectively, efficiently, and sustainably is among the most important undertakings in this century. While in the recent past, decision makers and social scientists faced significant constraints in obtaining the data needed to understand city dynamics and evaluate policies and practices, data are now abundant. Many cities have started to make a wide range of data sets available, see e.g., [24, 10, 7]. The challenge now is how to make sense of these data.

We examine one particularly important urban data set: taxi trips. In New York City, each day 13,000 taxis carry over one million passengers and make, on average, 500,000 trips—totaling over 170 million trips a year. Taxi trips are thus valuable sensors of city life. Consider the plot in Fig. 2, which shows how the number of trips per day varies

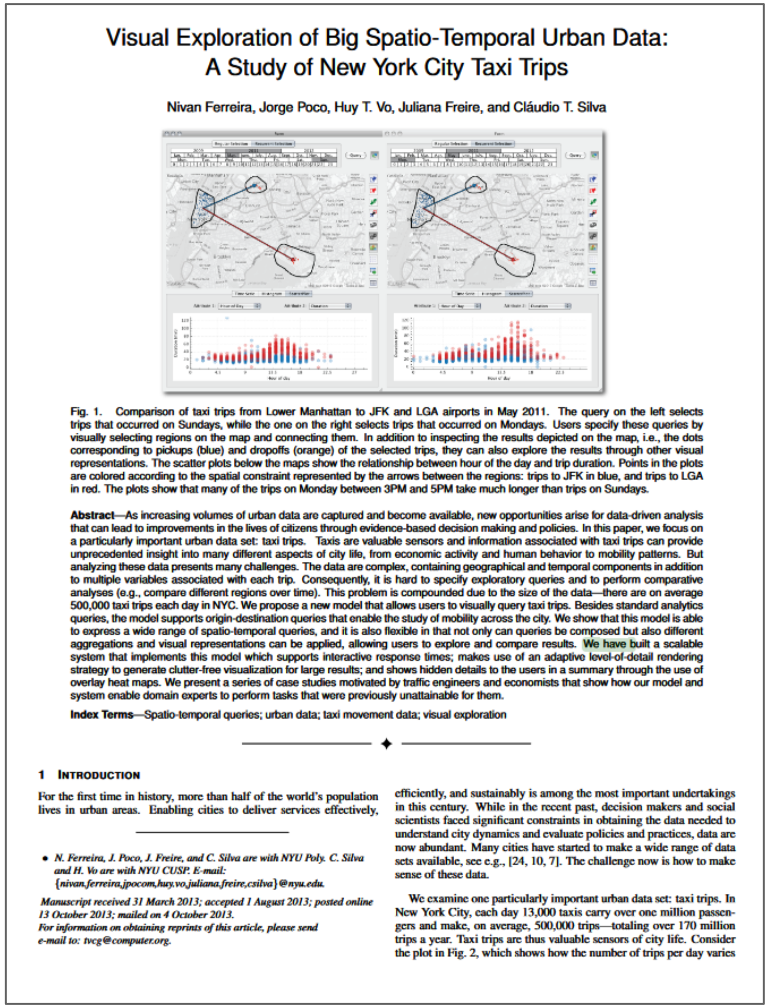
• *N. Ferreira, J. Poco, J. Freire, and C. Silva are with NYU Poly. C. Silva and H. Vo are with NYU CUSP. E-mail: {nivan.ferreira,jpocom,huy.vo,juliana.freire,csilva}@nyu.edu.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

[Ferreira et al., 2013]

TaxiVis

- Interaction capabilities that enable users to pose queries over all the dimensions of the data and flexibly explore the associated attributes.
- Compare spatiotemporal slices through multiple coordinated views.
- Interactively compose and refine queries, and generalize them by performing parameter sweeps.
- Efficient data stored, and adaptive level-of-detail rendering to provide clutter-free visualizations.



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

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

Fig. 1. Comparison of taxi trips from Lower Manhattan to JFK and LGA airports in May 2011. The query on the left selects trips that occurred on Sundays, while the one on the right selects trips that occurred on Mondays. Users specify these queries by visually selecting regions on the map and connecting them. In addition to inspecting the results depicted on the map, i.e., the dots corresponding to pickups (blue) and dropoffs (orange) of the selected trips, they can also explore the results through other visual representations. The scatter plots below the maps show the relationship between hour of the day and trip duration. Points in the plots are colored according to the spatial constraint represented by the arrows between the regions: trips to JFK in blue, and trips to LGA in red. The plots show that many of the trips on Monday between 3PM and 5PM take much longer than trips on Sundays.

Abstract—As increasing volumes of urban data are captured and become available, new opportunities arise for data-driven analysis that can lead to improvements in the lives of citizens through evidence-based decision making and policies. In this paper, we focus on a particularly important urban data set: taxi trips. Taxis are valuable sensors and information associated with taxi trips can provide unprecedented insight into many different aspects of city life, from economic activity and human behavior to mobility patterns. But analyzing these data presents many challenges. The data are complex, containing geographical and temporal components in addition to multiple variables associated with each trip. Consequently, it is hard to specify exploratory queries and to perform comparative analyses (e.g., compare different regions over time). This problem is compounded due to the size of the data—there are on average 500,000 taxi trips each day in NYC. We propose a new model that allows users to visually query taxi trips. Besides standard analytics queries, the model supports origin-destination queries that enable the study of mobility across the city. We show that this model is able to express a wide range of spatio-temporal queries, and it is also flexible in that not only can queries be composed but also different aggregations and visual representations can be applied, allowing users to explore and compare results. We have built a scalable system that implements this model which supports interactive response times; makes use of an adaptive level-of-detail rendering strategy to generate clutter-free visualization for large results; and shows hidden details to the users in a summary through the use of overlay heat maps. We present a series of case studies motivated by traffic engineers and economists that show how our model and system enable domain experts to perform tasks that were previously unattainable for them.

Index Terms—Spatio-temporal queries; urban data; taxi movement data; visual exploration

1 INTRODUCTION

For the first time in history, more than half of the world's population lives in urban areas. Enabling cities to deliver services effectively, efficiently, and sustainably is among the most important undertakings in this century. While in the recent past, decision makers and social scientists faced significant constraints in obtaining the data needed to understand city dynamics and evaluate policies and practices, data are now abundant. Many cities have started to make a wide range of data sets available, see e.g., [24, 10, 7]. The challenge now is how to make sense of these data.

We examine one particularly important urban data set: taxi trips. In New York City, each day 13,000 taxis carry over one million passengers and make, on average, 500,000 trips—totaling over 170 million trips a year. Taxi trips are thus valuable sensors of city life. Consider the plot in Fig. 2, which shows how the number of trips per day varies

- *N. Ferreira, J. Poco, J. Freire, and C. Silva are with NYU Poly. C. Silva and H. Vo are with NYU CUSP. E-mail: {nivan.ferreira,jpocom,huy.vo,juliana.freire,csilva}@nyu.edu.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

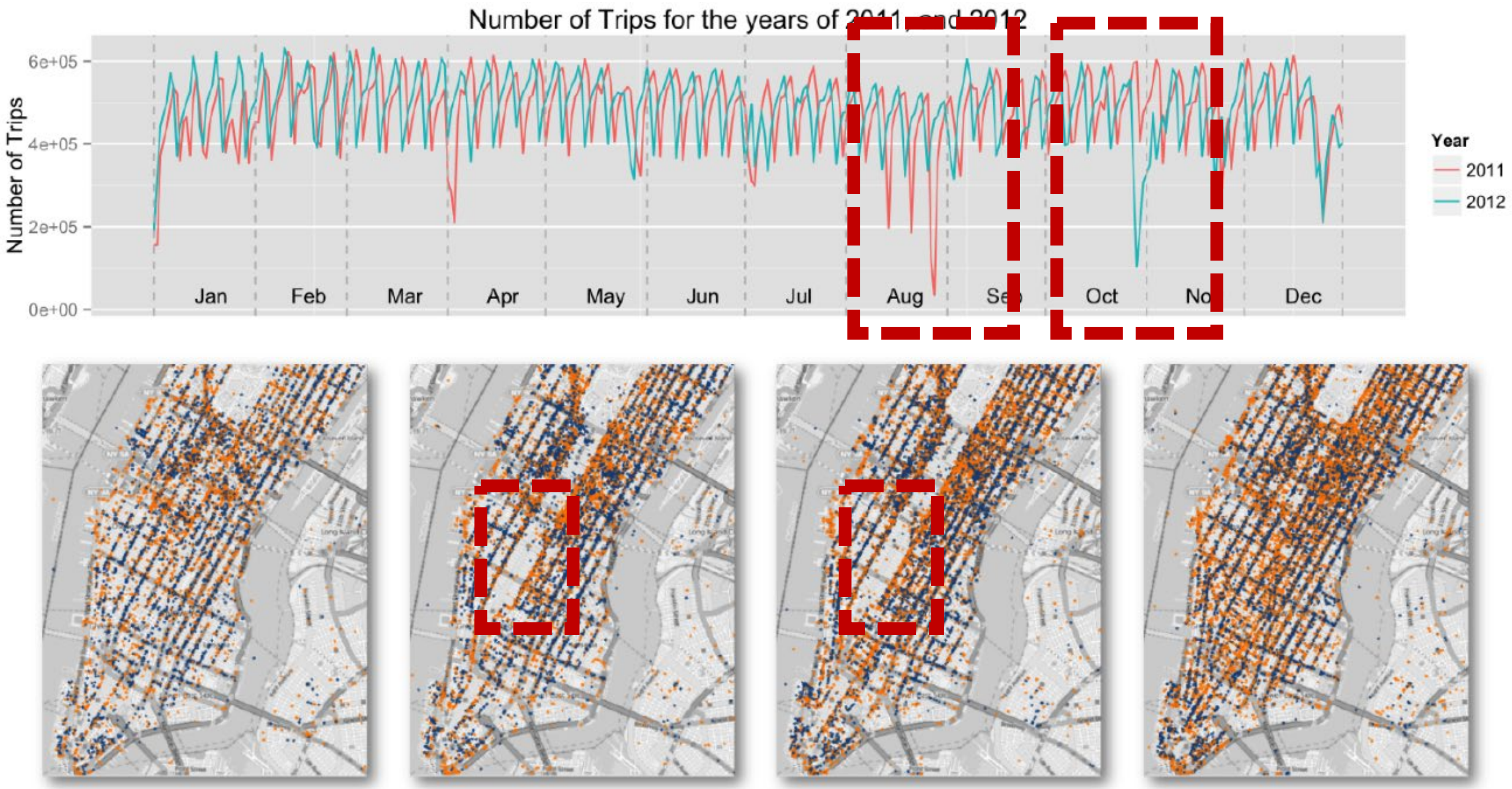
[Ferreira et al., 2013]

UIC COMPUTER SCIENCE

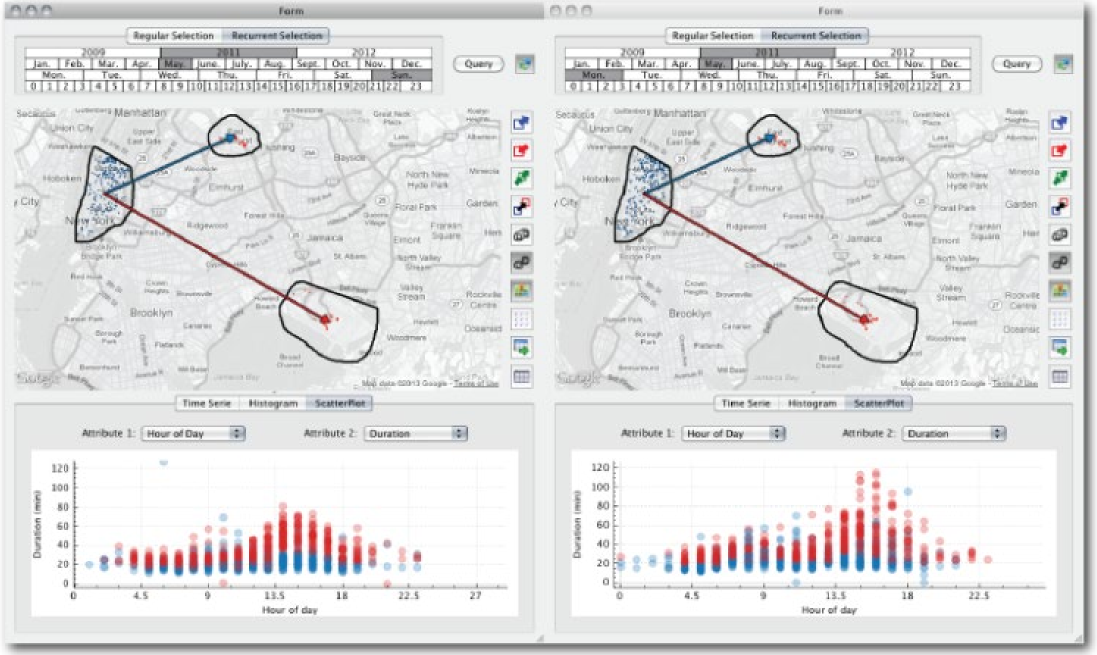
Taxi as sensors of city life

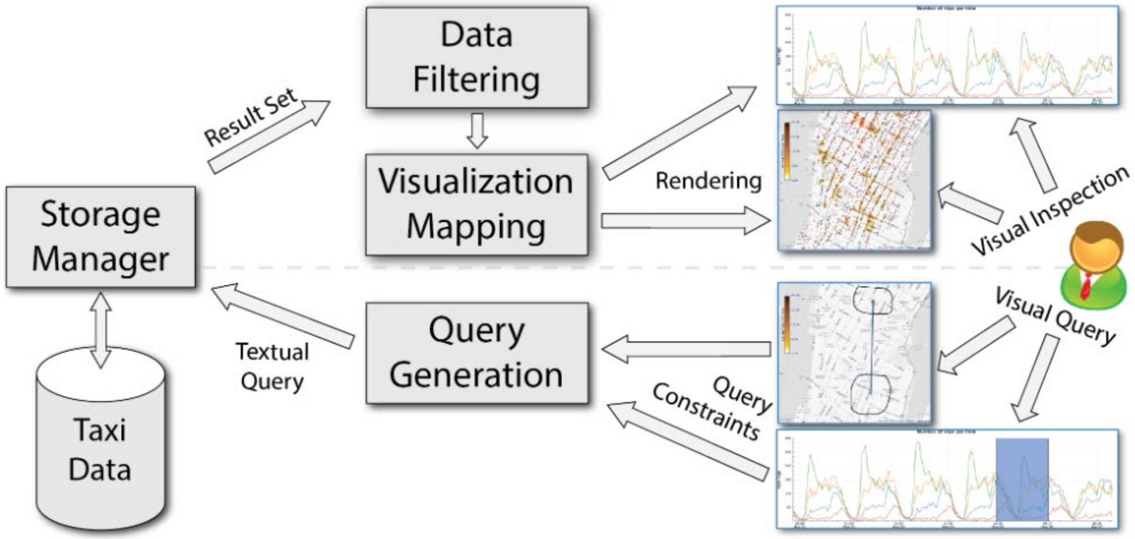
- Understanding the dynamics of the city, how different aspects of the data vary over space and time:
  - "*What is the average trip time from Midtown to the airports during weekdays?*"
  - "*How does the movement changes between Midtown and JFK throughout the day, over different days of the week?*"
  - "*How does the taxi fleet activity vary during weekdays?*"
- Ability to quickly test hypotheses: starting with one query about a specific place ("*movement patterns between Midtown and JFK*") and generalize to all neighborhoods.

Taxis as sensors of city life

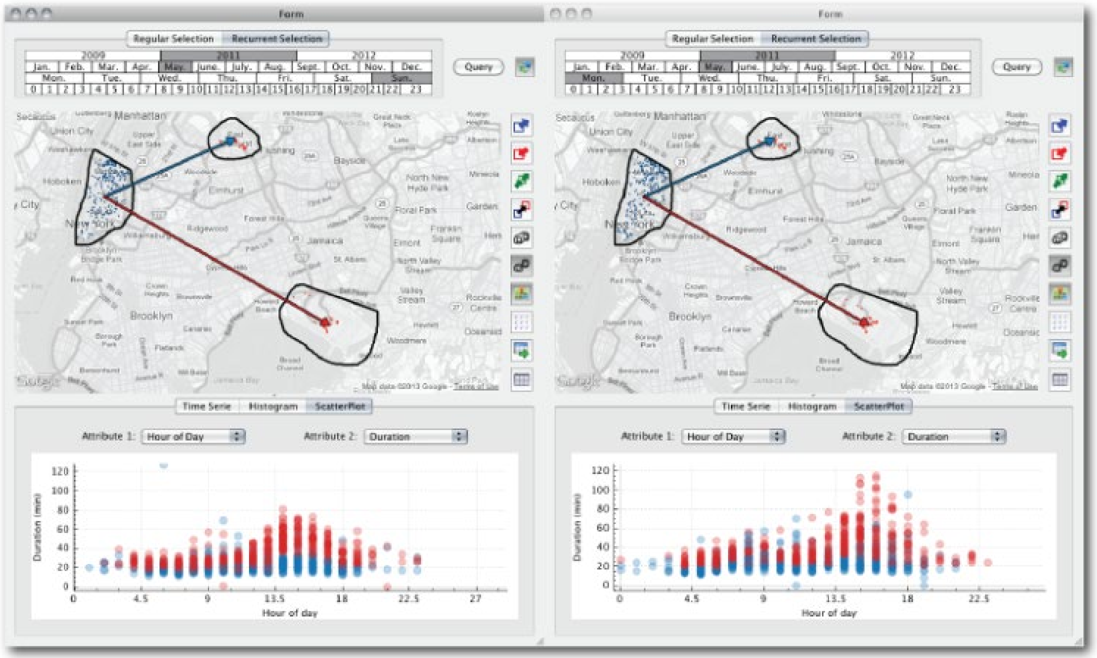


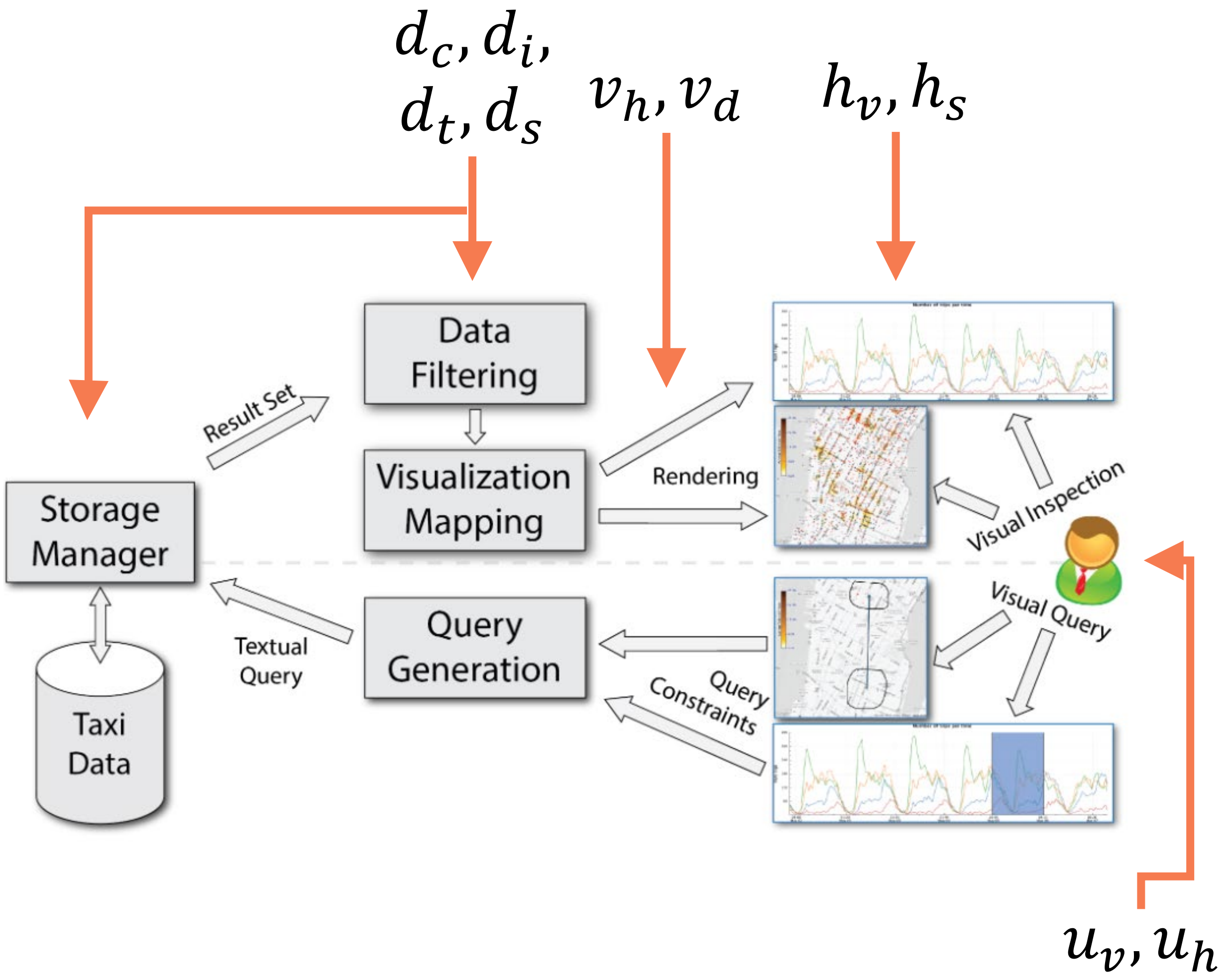
TaxiVis




Data Filtering
Visualization Mapping
Storage Manager
Taxi Data
Query Generation
Result Set
Rendering
Textual Query
Query Constraints
Visual Inspection
Visual Query

TaxiVis




$d_c, d_i,$
d_t, d_s
v_h, v_d
h_v, h_s
Result Set
Data Filtering
Visualization Mapping
Rendering
Storage Manager
Visual Inspection
Visual Query
Textual Query
Query Generation
Query Constraints
Taxi Data
u_v, u_h

Urban Mosaic: street-level images

" … a tool for exploring the urban fabric through a spatially and temporally dense data set of 7.7 million street-level images from New York City, captured over the period of a year. Working in collaboration with professional practitioners, we use Urban Mosaic to investigate questions of accessibility and mobility, and preservation and retrofitting."

Urban Mosaic: Visual Exploration of Streetscapes Using Large-Scale Image Data

Fabio Miranda[1], Maryam Hosseini[2], Marcos Lage[3], Harish Doraiswamy[1], Graham Dove[1], Cláudio T. Silva[1]
[1]New York University; [2]Rutgers University; [3]Universidade Federal Fluminense
[1]{fmiranda,harishd,grahamdove,csilva}@nyu.edu; [2]mary.hosseini@rutgers.edu; [3]mlage@ic.uff.br

ABSTRACT
Urban planning is increasingly data driven, yet the challenge of designing with data at a city scale and remaining sensitive to the impact at a human scale is as important today as it was for Jane Jacobs. We address this challenge with Urban Mosaic, a tool for exploring the urban fabric through a spatially and temporally dense data set of 7.7 million street-level images from New York City, captured over the period of a year. Working in collaboration with professional practitioners, we use Urban Mosaic to investigate questions of accessibility and mobility, and preservation and retrofitting. In doing so, we demonstrate how tools such as this might provide a bridge between the city and the street, by supporting activities such as visual comparison of geographically distant neighborhoods, and temporal analysis of unfolding urban development.

Author Keywords
Urban planning; Interactive visualization; Data analysis; Urban data

CCS Concepts
•**Human-centered computing → Human computer interaction (HCI);** *Visual analytics; Visualization toolkits;*

"A sense of place is built up, in the end, from many little things too, some so small people take them for granted, and yet the lack of them takes the flavor out of the city."

(Jane Jacobs, Downtown is for People)

INTRODUCTION
For those of us living in or visiting the world's major cities, their dynamism and complexity are immediately apparent. Yet urban planners and designers must work in a context where any single intervention, perhaps aimed at altering just one aspect, can have a wide ranging impact on a variety of interrelated components [12], affecting things at both a macro *city scale* and a micro *human scale*. Examples of changes at a city scale include, suburbanization, economic deconcentration, modification to transport infrastructure, rezoning and/or gentrification of neighborhoods, and major renewal projects. Examples of changes at a human scale, on the other hand, are reflected in the city's urban fabric, and include aspects that make a city livable, encourage walking, and contribute to the perception of safety; in other words affect the day to day lives of its inhabitants. This might be manifested in lighting, shadow, sky exposure, open-front shops, the details of building facades, etc. [32].

Responding to these challenges at the city scale, planners and designers have been aided by a rapid growth in data from urban environments, and so are able to turn to computational methods and large-scale data analysis, which increase understanding by quantifying different aspects of the city (e.g., [36, 13, 14, 29, 70, 67, 48]). However, while sensitivity to the impact of change at a human scale remains as important today as it was for Jane Jacobs and others in the 1950s and 1960s [47], analyses of suitable data, which emphasize qualitative, visual details, are often difficult and time-consuming to perform. Although there is an increasing availability of street-view images, which support a degree of virtual assessment and auditing of the built environment, their distribution is often temporally sparse and so analysis is limited.

This paper introduces Urban Mosaic, a tool for visually exploring the urban fabric. It responds to the challenges practitioners face by employing a newly available spatially and temporally dense data set of street-level images from New York City (NYC) Urban Mosaic is a visual exploration system designed to help practitioners in urban planning and design gain insight into the human scale impact of changes in the urban fabric. It utilizes state-of-the-art computer vision techniques for image similarity search and clustering, together with efficient spatio-temporal selection and aggregation over the image metadata to visually explore and map this image data set. It further allows the analysis to be augmented using spatio-temporal urban data from a variety of other sources (e.g., census, transport, crime, weather, housing market, zoning, noise complaints). The image data set used in this work contains 7.7 million images captured in the Manhattan and Brooklyn boroughs between April 2016 and April 2017 using car-mounted cameras. Urban Mosaic has been developed as a collaboration between researchers in urban planning, visual analytics, and HCI. We include a detailed

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376399

[Miranda et al., 2020]

Urban Mosaic: street-level images

- Demonstrates the potential for images from large-scale street-view data sets to help bridge data-driven urban planning and design at city scale and at the human level.
- Exploration of the urban fabric using large spatiotemporal collection of images.
- Visual comparison of geographically distant areas to analyze unfolding urban developments.

Urban Mosaic: Visual Exploration of Streetscapes Using Large-Scale Image Data

Fabio Miranda[1], Maryam Hosseini[2], Marcos Lage[3], Harish Doraiswamy[1], Graham Dove[1], Cláudio T. Silva[1]
[1]New York University; [2]Rutgers University; [3]Universidade Federal Fluminense
[1]{fmiranda,harishd,grahamdove,csilva}@nyu.edu; [2]mary.hosseini@rutgers.edu; [3]mlage@ic.uff.br

ABSTRACT
Urban planning is increasingly data driven, yet the challenge of designing with data at a city scale and remaining sensitive to the impact at a human scale is as important today as it was for Jane Jacobs. We address this challenge with Urban Mosaic, a tool for exploring the urban fabric through a spatially and temporally dense data set of 7.7 million street-level images from New York City, captured over the period of a year. Working in collaboration with professional practitioners, we use Urban Mosaic to investigate questions of accessibility and mobility, and preservation and retrofitting. In doing so, we demonstrate how tools such as this might provide a bridge between the city and the street, by supporting activities such as visual comparison of geographically distant neighborhoods, and temporal analysis of unfolding urban development.

Author Keywords
Urban planning; Interactive visualization; Data analysis; Urban data

CCS Concepts
•**Human-centered computing → Human computer interaction (HCI);** *Visual analytics; Visualization toolkits;*

"A sense of place is built up, in the end, from many little things too, some so small people take them for granted, and yet the lack of them takes the flavor out of the city."

(Jane Jacobs, Downtown is for People)

INTRODUCTION
For those of us living in or visiting the world's major cities, their dynamism and complexity are immediately apparent. Yet urban planners and designers must work in a context where any single intervention, perhaps aimed at altering just one aspect, can have a wide ranging impact on a variety of interrelated components [12], affecting things at both a macro *city scale* and a micro *human scale*. Examples of changes at a city scale include, suburbanization, economic deconcentration, modification to transport infrastructure, rezoning and/or gentrification of neighborhoods, and major renewal projects. Examples of changes at a human scale, on the other hand, are reflected in the city's urban fabric, and include aspects that make a city livable, encourage walking, and contribute to the perception of safety; in other words affect the day to day lives of its inhabitants. This might be manifested in lighting, shadow, sky exposure, open-front shops, the details of building facades, etc. [32].

Responding to these challenges at the city scale, planners and designers have been aided by a rapid growth in data from urban environments, and so are able to turn to computational methods and large-scale data analysis, which increase understanding by quantifying different aspects of the city (e.g., [36, 13, 14, 29, 70, 67, 48]). However, while sensitivity to the impact of change at a human scale remains as important today as it was for Jane Jacobs and others in the 1950s and 1960s [47], analyses of suitable data, which emphasize qualitative, visual details, are often difficult and time-consuming to perform. Although there is an increasing availability of street-view images, which support a degree of virtual assessment and auditing of the built environment, their distribution is often temporally sparse and so analysis is limited.

This paper introduces Urban Mosaic, a tool for visually exploring the urban fabric. It responds to the challenges practitioners face by employing a newly available spatially and temporally dense data set of street-level images from New York City (NYC) Urban Mosaic is a visual exploration system designed to help practitioners in urban planning and design gain insight into the human scale impact of changes in the urban fabric. It utilizes state-of-the-art computer vision techniques for image similarity search and clustering, together with efficient spatio-temporal selection and aggregation over the image metadata to visually explore and map this image data set. It further allows the analysis to be augmented using spatio-temporal urban data from a variety of other sources (e.g., census, transport, crime, weather, housing market, zoning, noise complaints). The image data set used in this work contains 7.7 million images captured in the Manhattan and Brooklyn boroughs between April 2016 and April 2017 using car-mounted cameras. Urban Mosaic has been developed as a collaboration between researchers in urban planning, visual analytics, and HCI. We include a detailed

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376399

[Miranda et al., 2020]

Temporally-dense street-level images

Objectives

Support the interactive analysis of the city at the micro scale, over geographically distant regions.

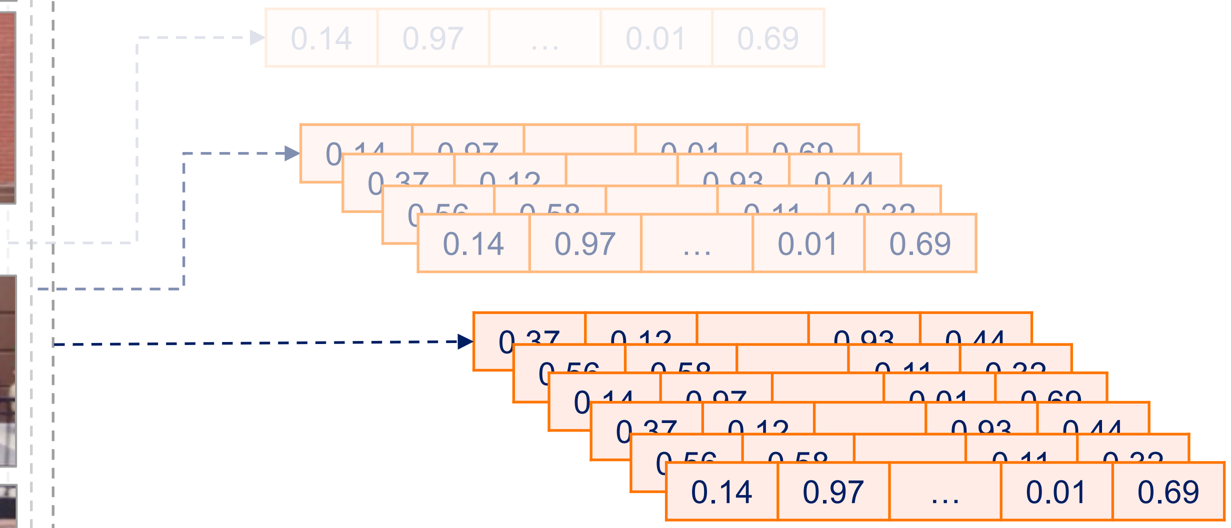
Comparison of the urban fabric in different regions

Assessment of features in the built environment

Assessment of walkability and accessibility

Image query composition

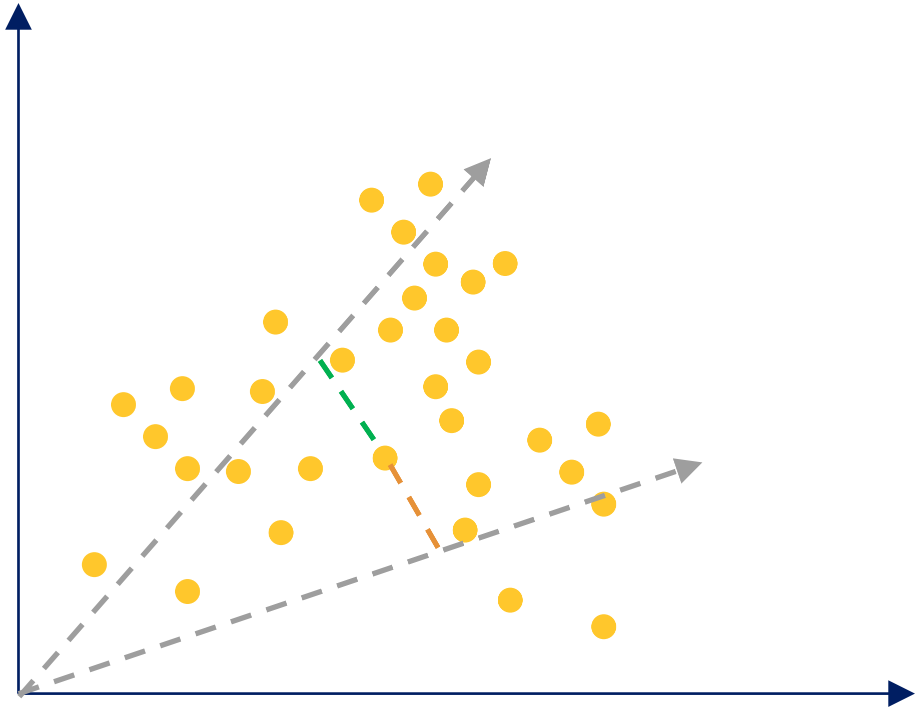
Image embeddings



21 embeddings per image

2.5 TB worth of embeddings data

Locality sensitive hashing



0.37	0.12	…	0.93	0.44

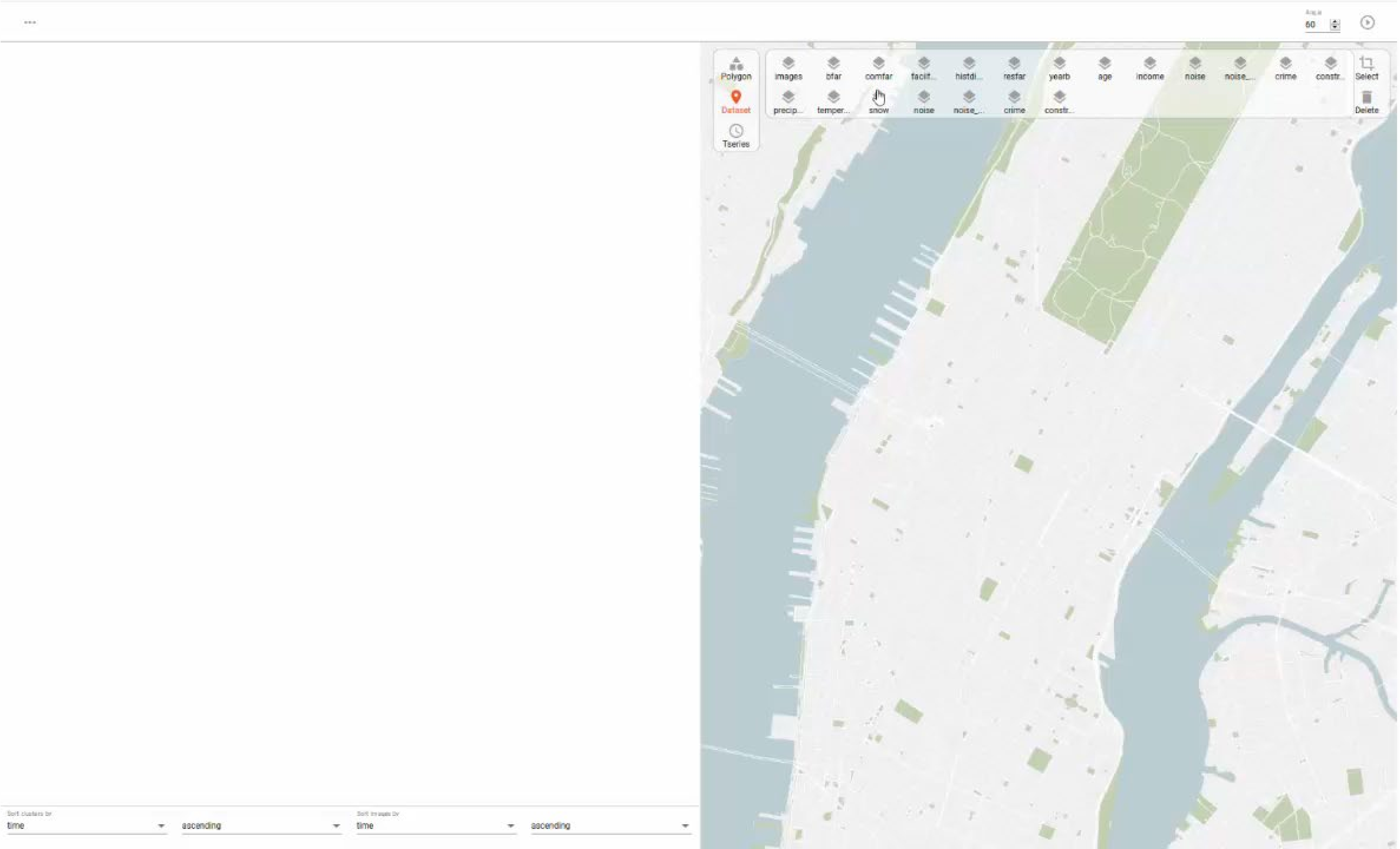
4 kbytes per embedding

0	0	…	1	0

64 bytes per embedding

$$\alpha_{1,2} = \cos^{-1}\left(\frac{\overrightarrow{v_1} \cdot \overrightarrow{v_2}}{|\overrightarrow{v_1}||\overrightarrow{v_2}|}\right)$$

21 GB worth of embeddings data

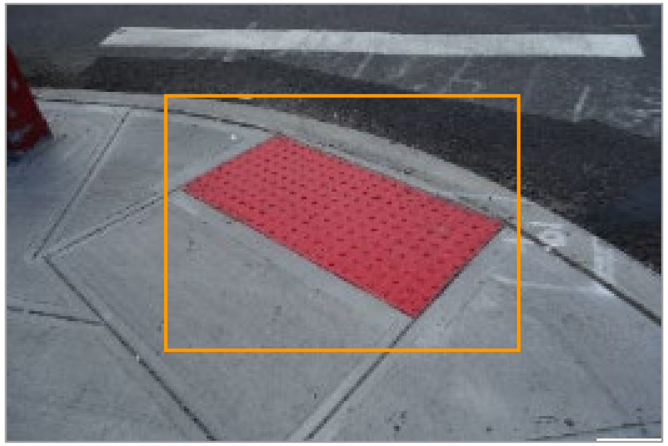

60
Polygon
Dataset
Tseries
images
bfar
comfar
facit...
histdi...
resfar
yearb
age
income
noise
noise_...
crime
constr...
Select
precip...
temper...
snow
noise
noise_...
crime
constr...
Delete
Sort clusters by
time
ascending
Sort images by
time
ascending

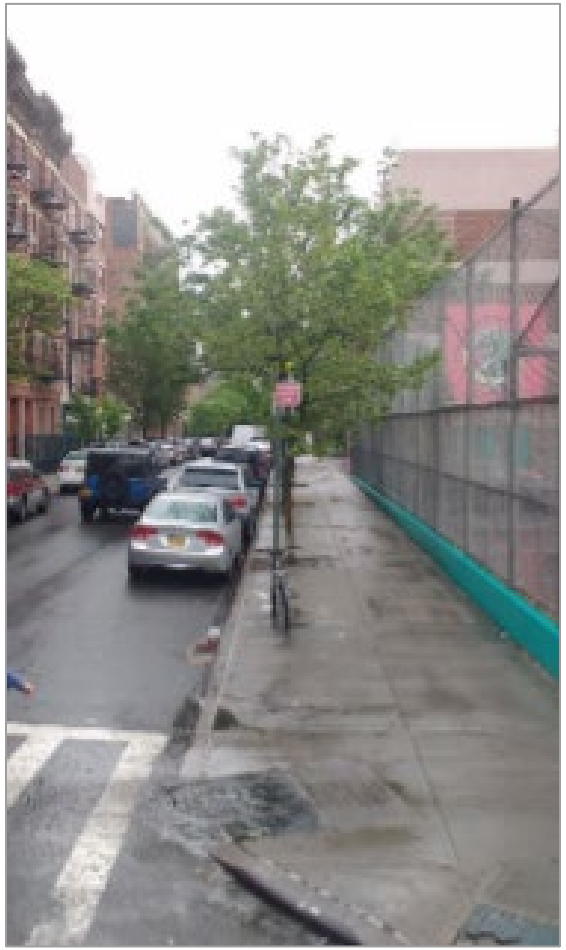
Accessibility for older adults

Focuses on interventions that enable older adults to “Age in Place” and prevent outdoor fall.

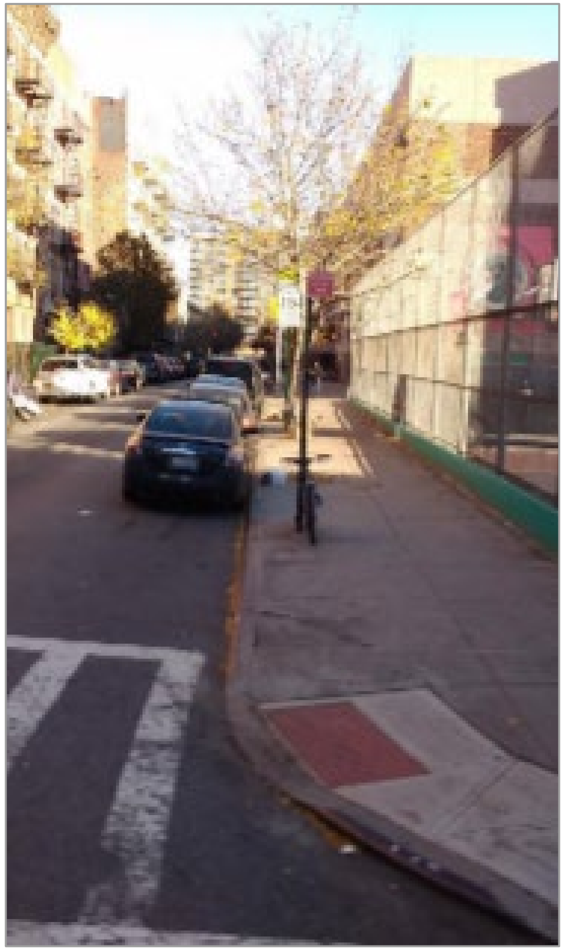
Impact of the neighborhood environment on falls.

Accessibility: installation of tactile pavings

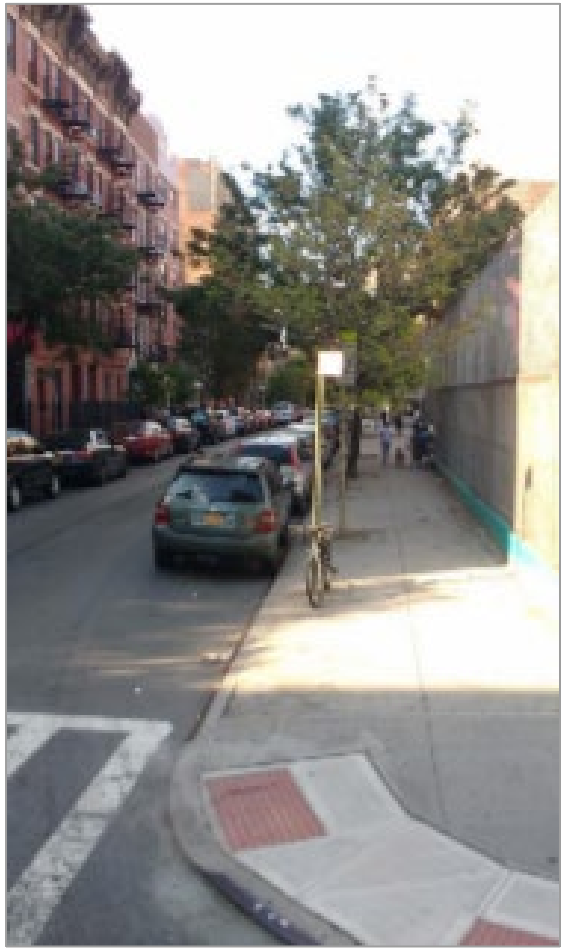




June

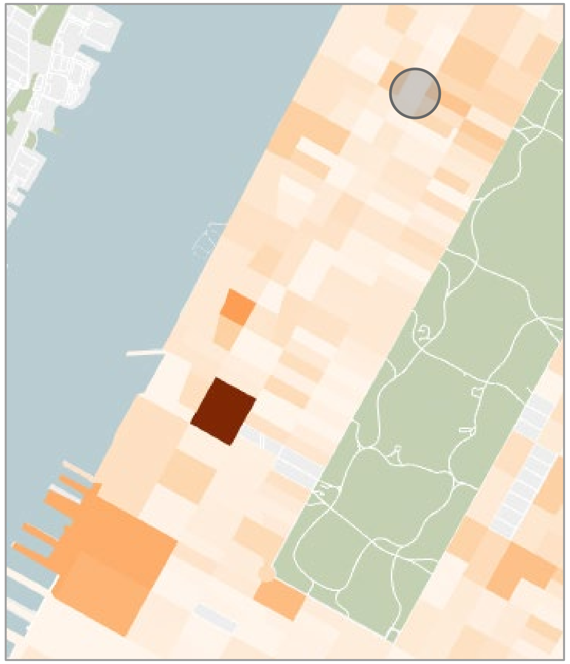


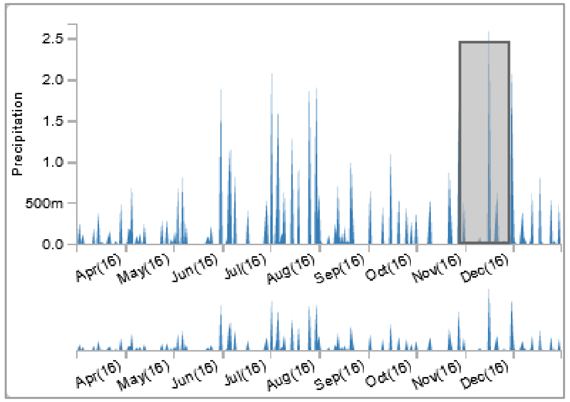
August

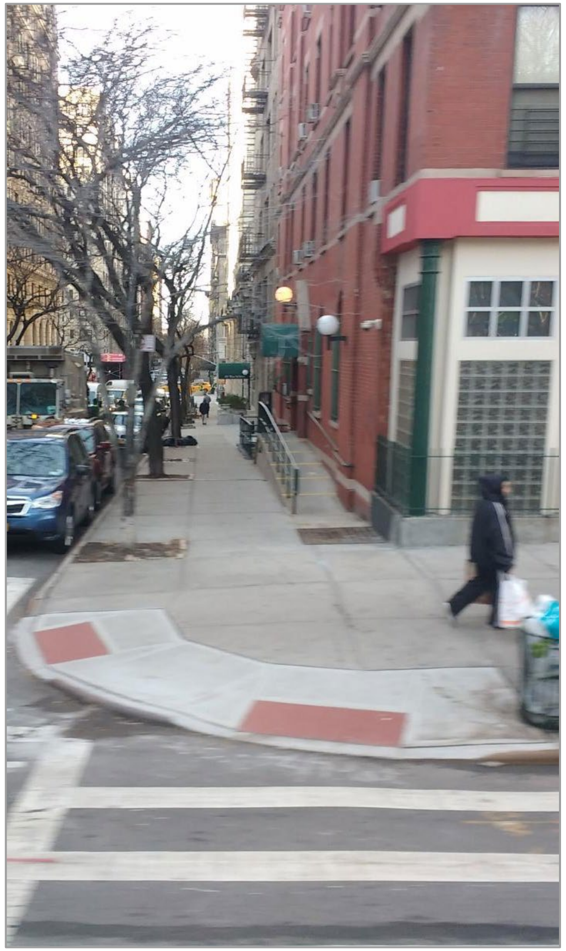


October

Accessibility: assessing hazards for older adults



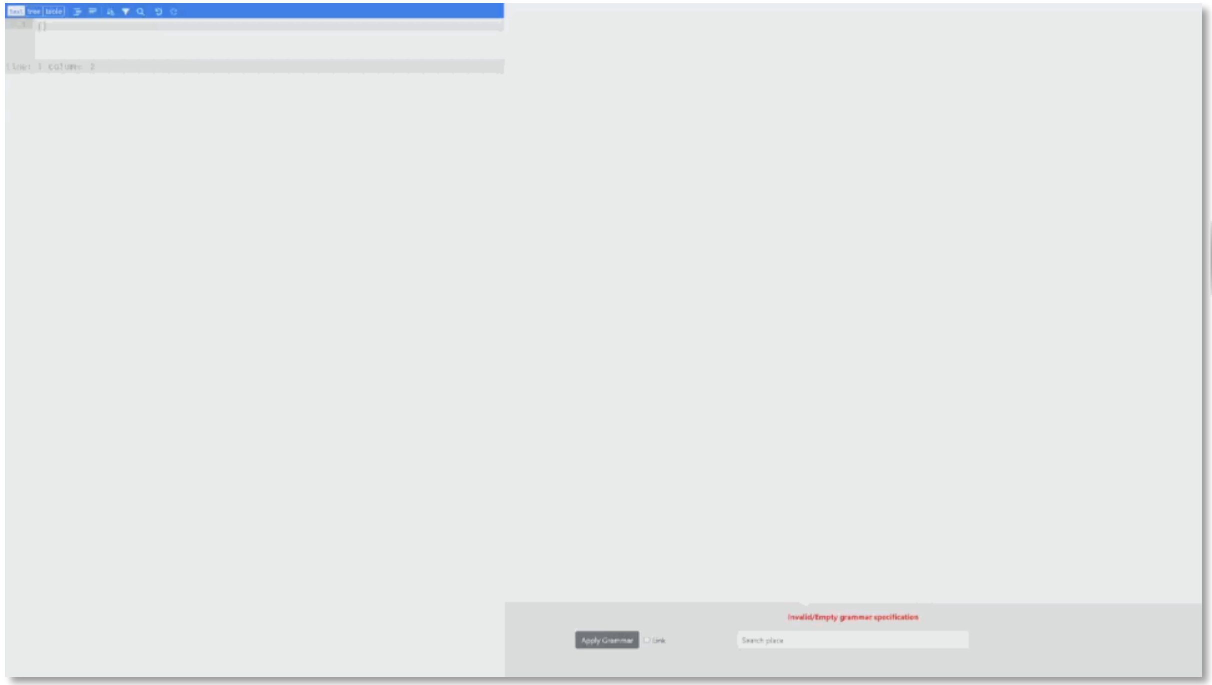




Main takeaways

- General approaches might not scale.
- Data visual analytics might require structures specifically designed for a given task.
- Understand the data, and how it can be transformed to best fit different computational resources and architectures.

The Urban Toolkit



UrbanTK

Visual Analysis of Uncertainty in Trajectories [Lu et al., 2014]

TrajGraph [Huang et al., 2015]

SemanticTraj [Al-Dohuki et al., 2016]

Spatiotemporal Statistical Data [Kim et al., 2017]

Forecasting Road Traffic Congestion [Lee et al., 2019]

A study of new york city taxi trips [Ferreira et al., 2013]

MaraVis [Li et al., 2020]

SEEVis [Li et al., 2020]

Air Pollution Problem in Hong Kong [Qu et al., 2007]

Air Temperature [Gautier et al., 2020]

Urban Mosaic [Miranda et al., 2020]

Location2vec [Zhu et al., 2019]

A correlative analysis process in a visual analytics environment [Malik et al., 2012]

HydroQual [Accorsi et al., 2014]

Shadow accrual maps [Miranda et al., 2018]

Urban Pulse [Miranda et al., 2016]

Urban visual analytics surveys

Visual Analytics in Urban Computing: An Overview

Yixian Zheng, Wenchao Wu, Yuanzhe Chen, Huamin Qu, *Member, IEEE*, and Lionel M. Ni, *Fellow, IEEE*

Abstract—Nowadays, various data collected in urban context provide unprecedented opportunities for building a smarter city through urban computing. However, due to heterogeneity, high complexity and large volumes of these urban data, analyzing them is not an easy task, which often requires integrating human perception in analytical process, triggering a broad use of visualization. In this survey, we first summarize frequently used data types in urban visual analytics, and then elaborate on existing visualization techniques for time, locations and other properties of urban data. Furthermore, we discuss how visualization can be combined with automated analytical approaches. Existing work on urban visual analytics is categorized into two classes based on different outputs of such combinations: 1) For *data exploration and pattern interpretation*, we describe representative visual analytics tools designed for better insights of different types of urban data. 2) For *visual learning*, we discuss how visualization can help in three major steps of automated analytical approaches (i.e., cohort construction; feature selection & model construction; result evaluation & tuning) for a more effective machine learning or data mining process, leading to sort of artificial intelligence, such as a classifier, a predictor or a regression model. Finally, we outlook the future of urban visual analytics, and conclude the survey with potential research directions.

Index Terms—Urban computing, visual analytics, visualization, visual learning, spatio-temporal, multivariate

1 Introduction

With the development of science and technology, urbanization process has been accelerating worldwide, which on one hand improves people's life quality, on the other hand gives rise to serious problems, such as environmental pollution, traffic congestion and ever-increasing quite a few issues which have not been addressed satisfactorily. Recently, Zheng et al. [3] presented a survey on urban computing, which introduced general framework, key research problems, methodologies, and applications mainly based on automated data mining approaches. However, as

> 150 papers (Zheng et al., 2016)

A survey of urban visual analytics: Advances and future directions

Zikun Deng[1], Di Weng[2] (✉), Shuhan Liu[1], Yuan Tian[1], Mingliang Xu[3,4], and Yingcai Wu[1] (✉)

© The Author(s) 2022.

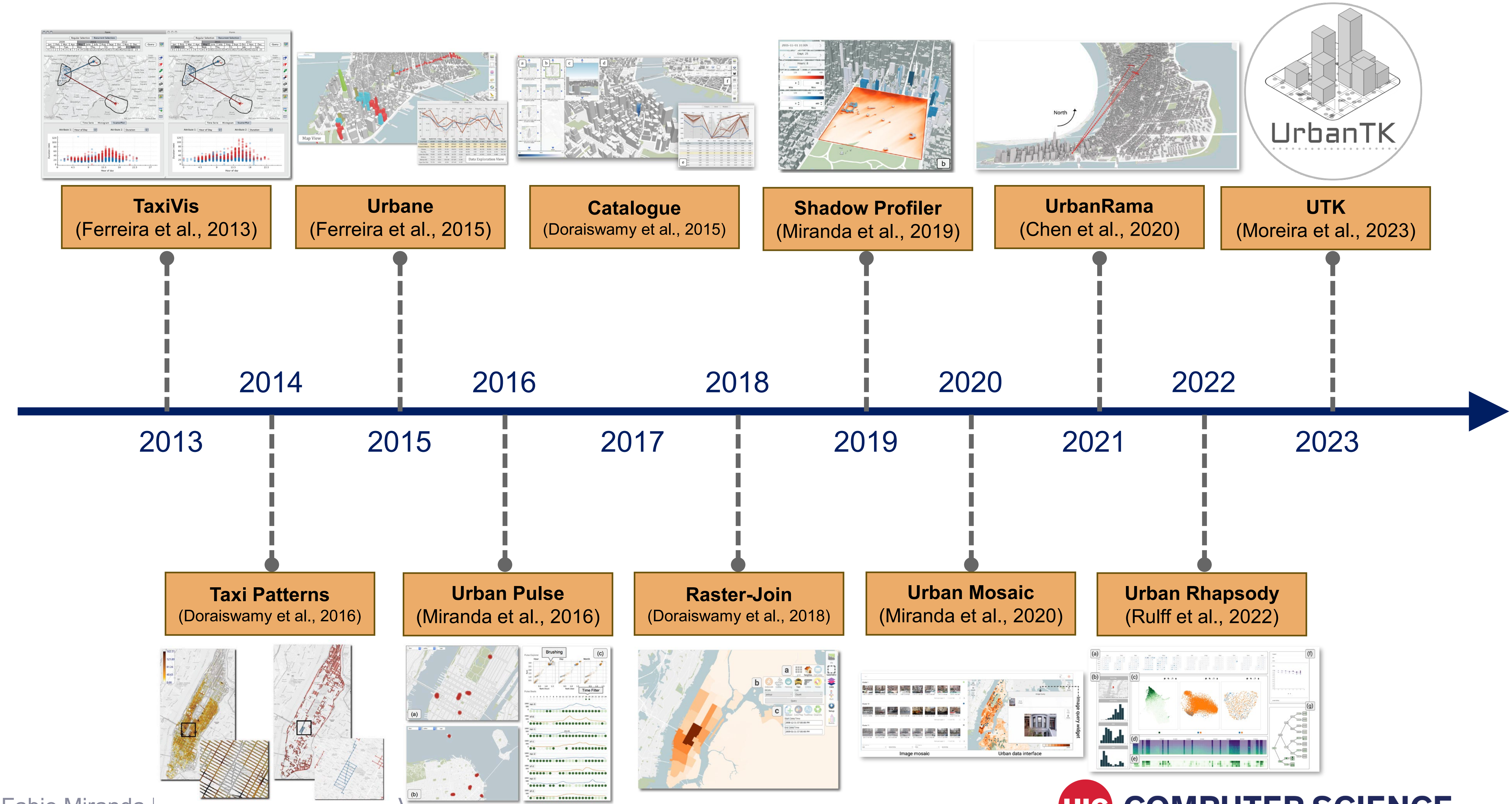
Abstract Developing effective visual analytics systems demands care in characterization of domain problems and integration of visualization techniques and computational models. Urban visual analytics has already achieved remarkable success in tackling urban problems and providing fundamental services for smart cities. To promote further academic research and assist the development of industrial urban analytics systems, we comprehensively review urban visual analytics studies from four perspectives. In particular, we identify 8 urban domains and 22 types of popular visualization, analyze 7 types of computational method, and categorize existing systems into 4 types based

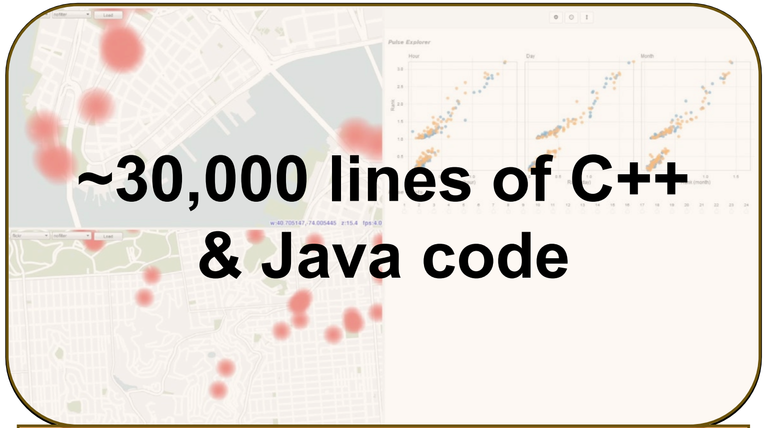
knowledge and expertise into the analysis loop. Thus, urban visual analytics [7] is used to empower urban experts using a combination of intuitive data visualization and fast computational methods, enabling experts to visually and interactively perceive, explore, manipulate, and reason about urban data [8].

When developing an urban visual analytics approach, practitioners like urban analysts and researchers may have the following four questions:

1. Which urban *domain problems* have been solved or remain unsolved by visual analytics?
2. What *visualization* techniques have been applied to visually interpret urban data?

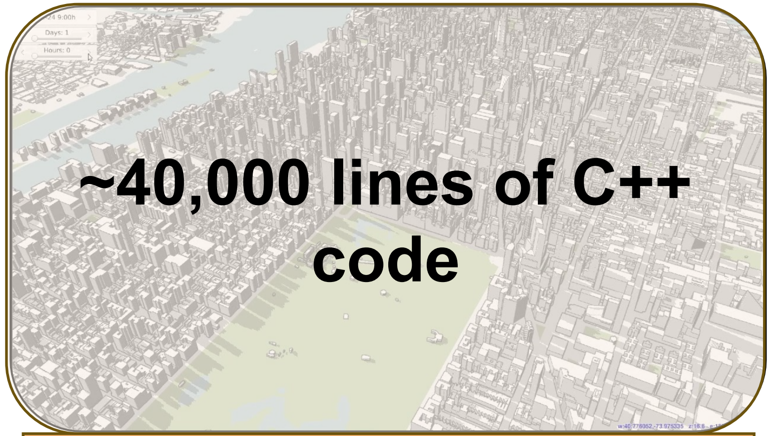
> 200 papers (Deng et al., 2022)


TaxiVis
(Ferreira et al., 2013)
Urbane
(Ferreira et al., 2015)
Catalogue
(Doraiswamy et al., 2015)
Shadow Profiler
(Miranda et al., 2019)
UrbanRama
(Chen et al., 2020)
UTK
(Moreira et al., 2023)
UrbanTK
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Taxi Patterns
(Doraiswamy et al., 2016)
Urban Pulse
(Miranda et al., 2016)
Raster-Join
(Doraiswamy et al., 2018)
Urban Mosaic
(Miranda et al., 2020)
Urban Rhapsody
(Rulff et al., 2022)
Map View
Data Exploration View
North
Brushing
Time Filter
Image mosaic
Urban data interface



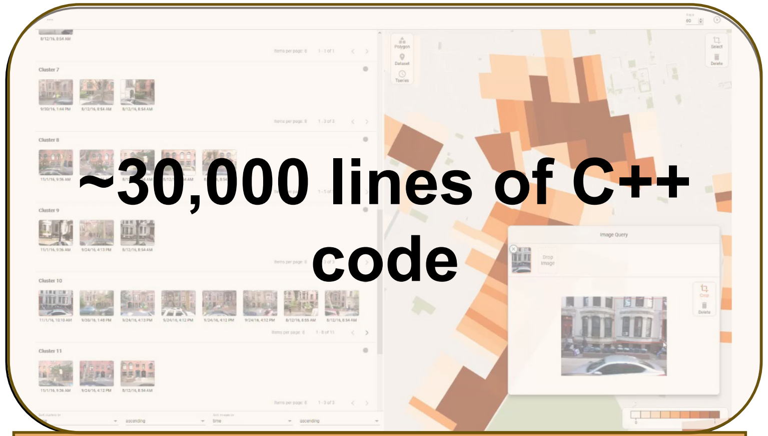
~30,000 lines of C++ & Java code

Urban Pulse:
Large-scale data mining of social media data



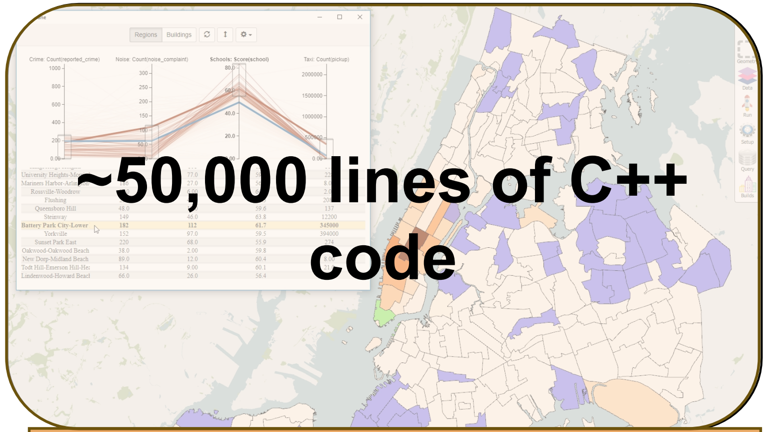
~40,000 lines of C++ code

Shadow Profiler:
City-scale assessment of sunlight access



~30,000 lines of C++ code

Urban Mosaic:
Interactive exploration of large imagery data



~50,000 lines of C++ code

Urbane:
Interactive exploration of large data

Current state of urban tools

Lack of flexibility:
Tools and techniques are not translatable to other domains or regions
(Acuto et al., 2018)

Lack of extensibility:
Hard to add new functionalities needed for specific workflows
(Lobo et al., 2020)

Lack of reproducibility:
Results are rarely reproducible
(Ziegler and Chasins, 2023)

Lack of scalability:
Analyses are rarely interactive, limiting the number of hypothesis from the exploratory process
(Ziegler and Chasins, 2023)

Lack of accessibility:
Tools and techniques are often difficult to use, limiting stakeholders' ability to conduct large-scale analysis
(Kontokosta, 2021)

How to support (1) accessible, (2) flexible, (3) extensible, (4) reproducible and (5) scalable tools across domains and users?

```cpp
#include "UrbaneMapView.hpp"

#include <QApplication>
#include "../MapView/BuildingRenderingLayer.hpp"

#include "../MassingGeneration/massinggeneration.h"
#include "../Util/ColorMapDivergent.hpp"
#include "UrbaneManager.hpp"

#include <QElapsedTimer>
#include <QThread>
#include <QDir>

#include <vector>

UrbaneMapView::UrbaneMapView(const QString &filename, const QRectF &vp, QWidget *parent)
    : MapView(filename,vp,parent), graphLayer(NULL)
{
    initialized = false;
    skyExposureData = false;
    this->centerIndex = GridIndex(1024,1024);
    this->currentLayer = NULL;
    this->lotUpdate = true;
}

UrbaneMapView::~UrbaneMapView() {}

void UrbaneMapView::initializeGL() {
    if(!initialized) {
        MapView::initializeGL();
        this->buildingScore.initComputeShader();
        this->skyScore.initComputeShader();
    }
    initialized = true;
}

void UrbaneMapView::paintGL()
{
    this->showOsd(false);

    // Lot data initialization in manager
    // TODO Don't know of a better place to do this
    if(lotUpdate && this->parcelLayer->isDataReady()) {
        updateLotDataDB();
        lotUpdate = false;
    }

    UrbaneManager *manager = UrbaneManager::getInstance();
    QPair<RenderingOperation, UIOperation> state = manager->getState();

    RenderingOperation operation = state.first;
    UIOperation what = state.second;
    switch(operation) {
    case RenderingOperation::UpdateVis:
    {
        bool updateFunction = false;
```

Abstracts low-level functionalities

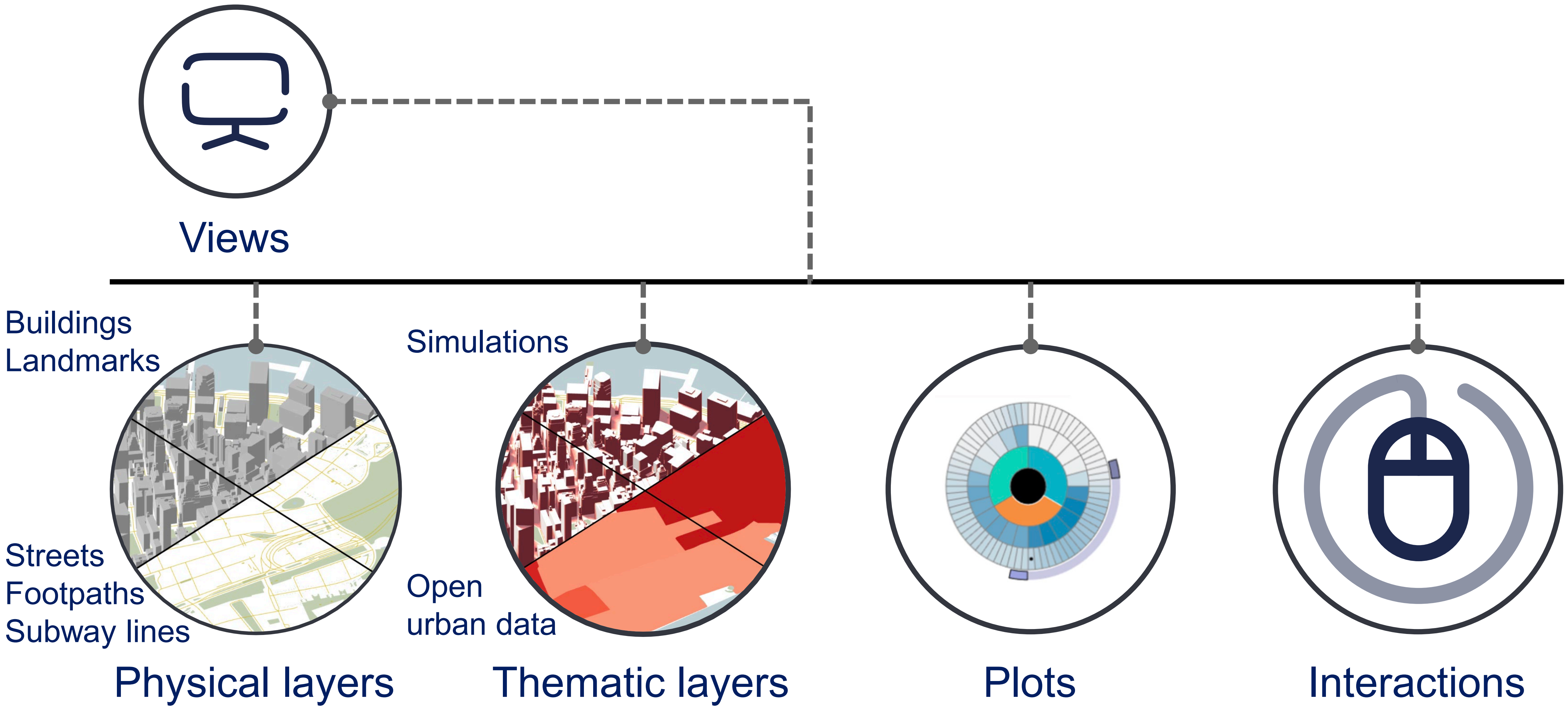
Easy access to data analytics

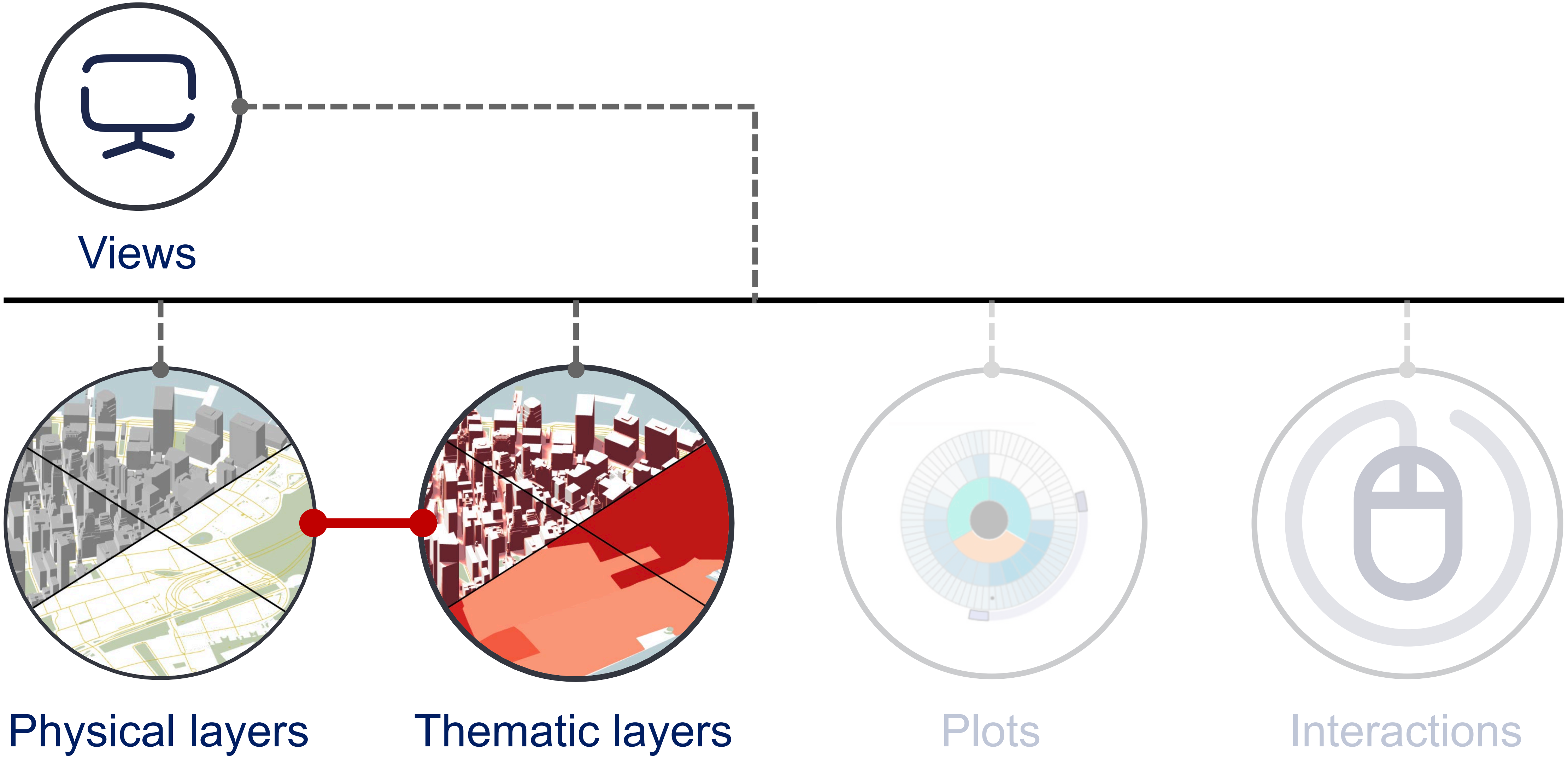
Self-contained & sharable JSON file

Lower the barrier for the construction of urban tools & systems

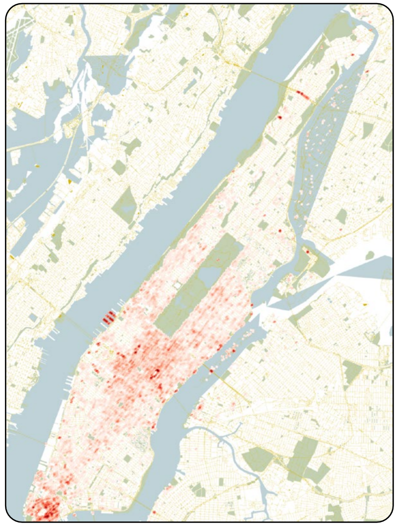
Views

```
{
  “views": [
    {
      "map": {
        "camera": {
          "position": […],
          "direction": {…}
        },
        …
      },
      "plots": […],
      "knots": […]
    }
  ]
}
```

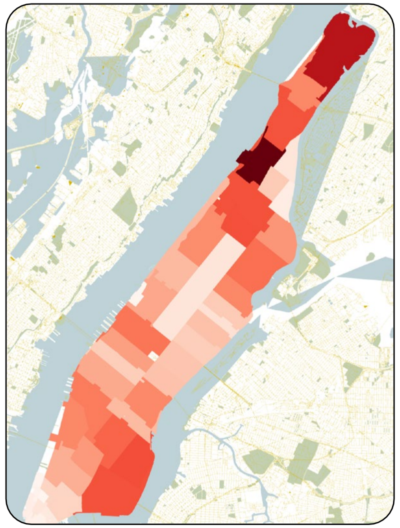

Views
Buildings
Landmarks
Streets
Footpaths
Subway lines
Simulations
Open
urban data
Physical layers
Thematic layers
Plots
Interactions


Views
Physical layers
Thematic layers
Plots
Interactions

Views

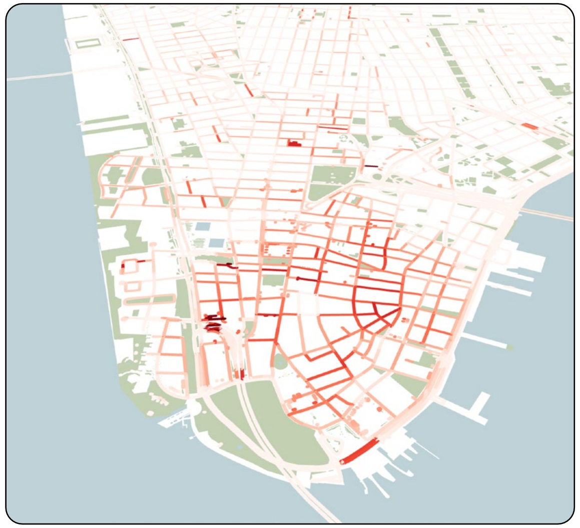


Grid

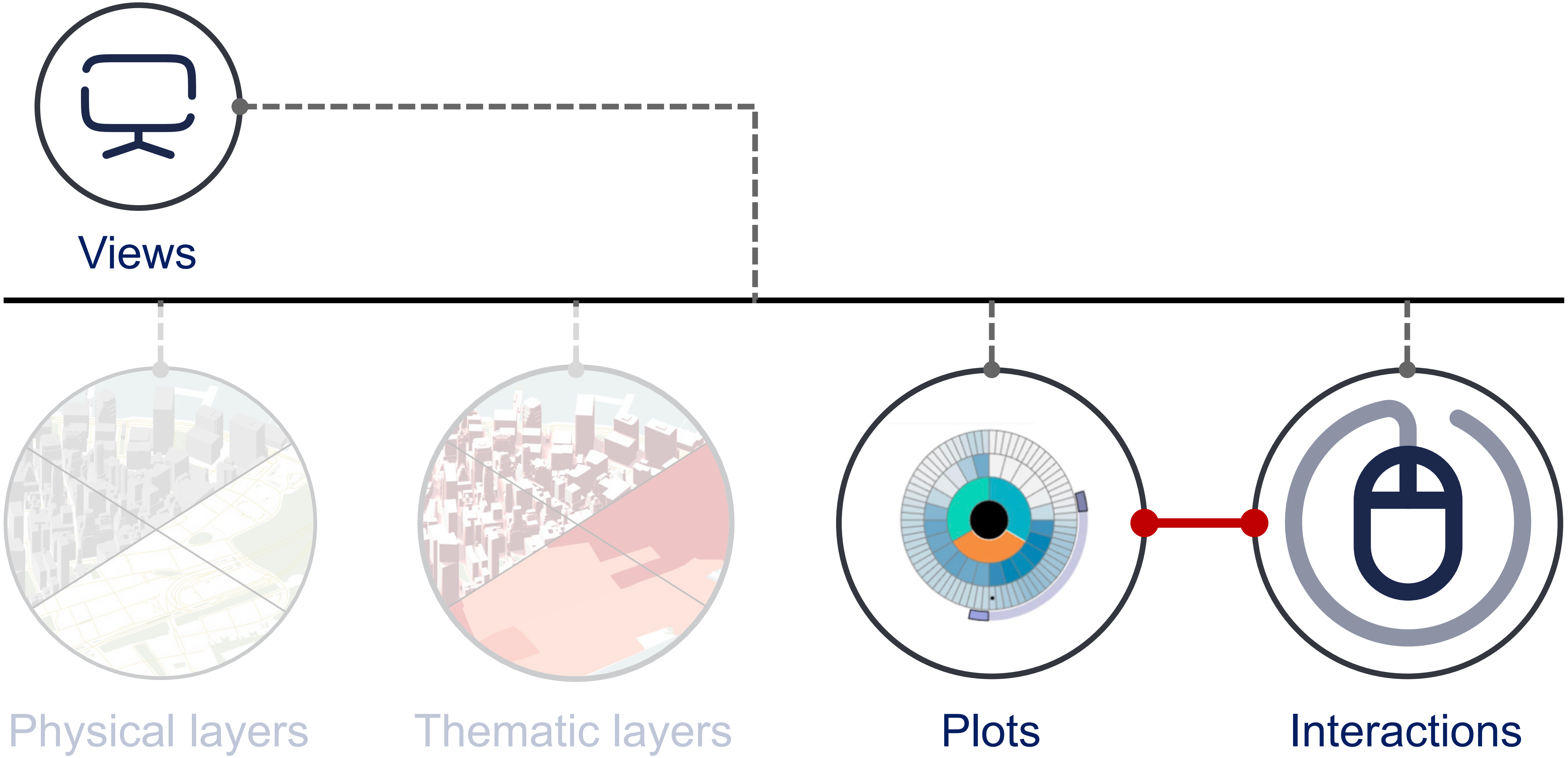


Areas

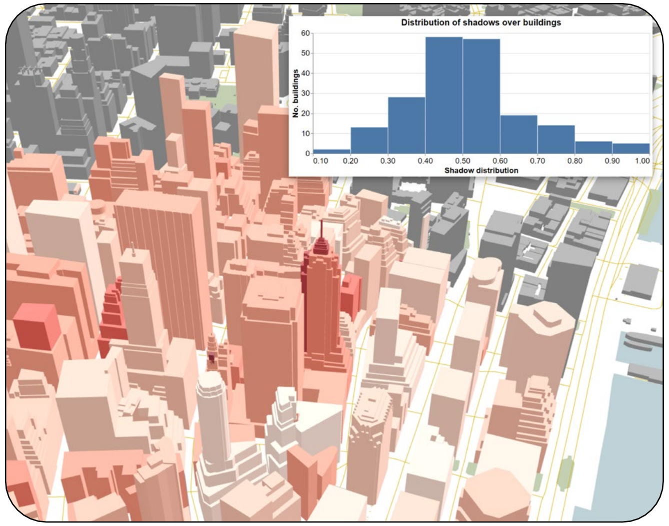
Buildings



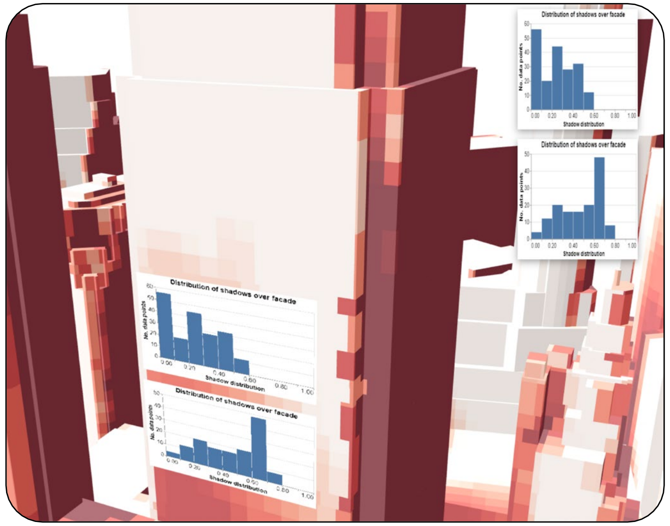
Networks


Views
Physical layers
Thematic layers
Plots
Interactions

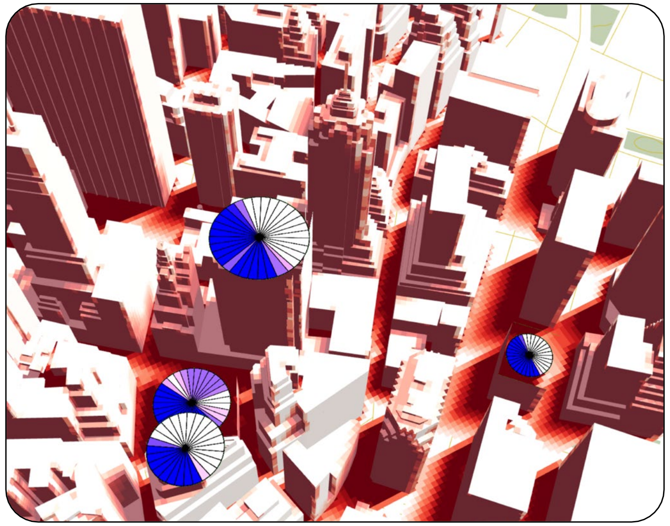
Views



Juxtaposed



Embedded (surface)

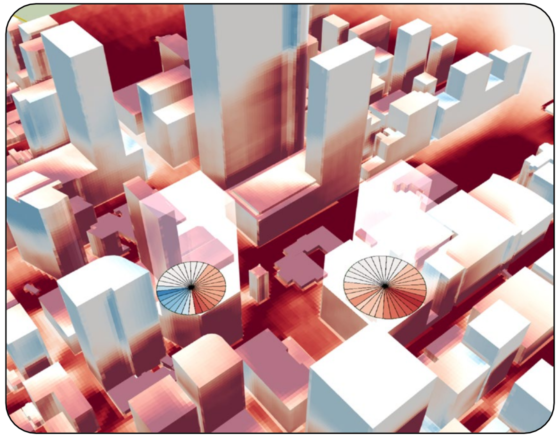


Embedded (crosscut)

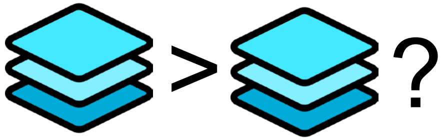
Taxonomy by Mota et al. (VIS 2022)

Operating with layers

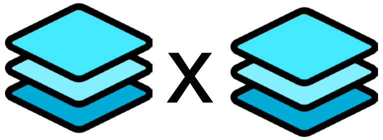
Arithmetic operations

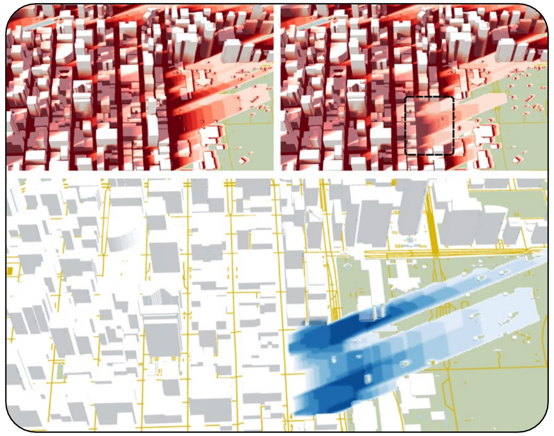


Logical operations

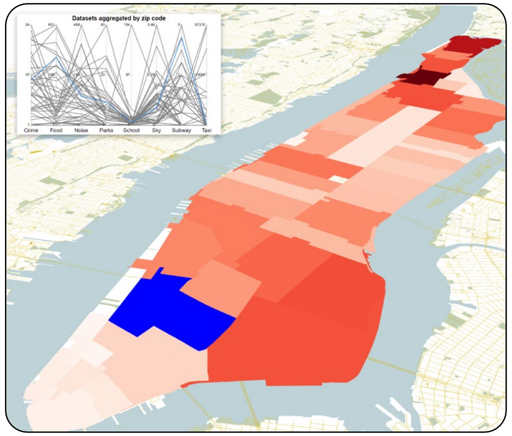


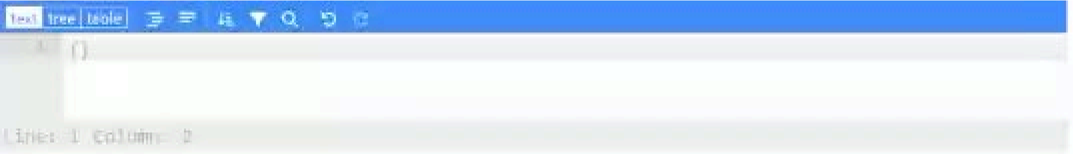
What-if analyses





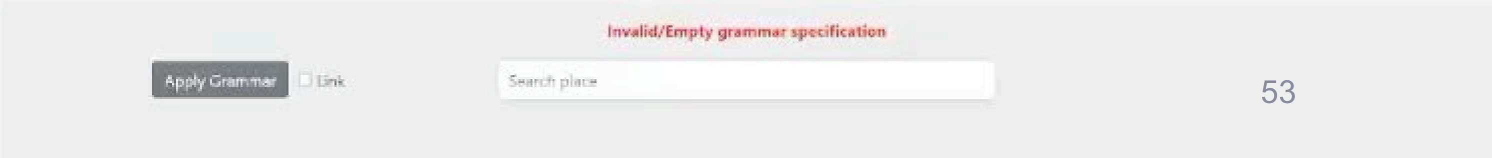
Data exploration





Editing the JSON specification

Visualization



Data functionalities

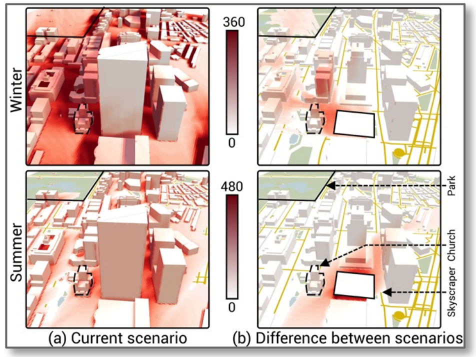
- Converting data: Pandas’ DataFrame, GeoJSON, shapefile, CSV and Protocolbuffer Binary Format (PBF).
- Loading data: OpenStreetMap (OSM).
- Generating data: Sunlight access simulation (Miranda et al., TVCG 2018).

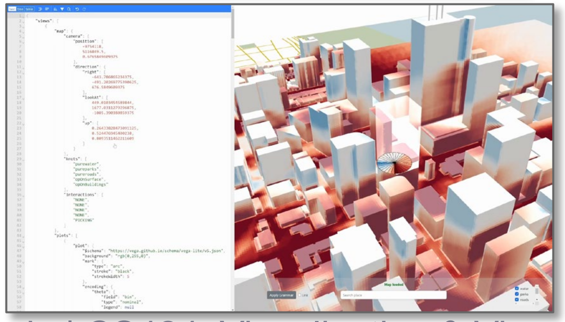
```
import utk
uc = utk.OSM.load(‘Chicago,USA’, layers=[‘buildings’])
uc.save(‘chicago’)
uc.view()
```

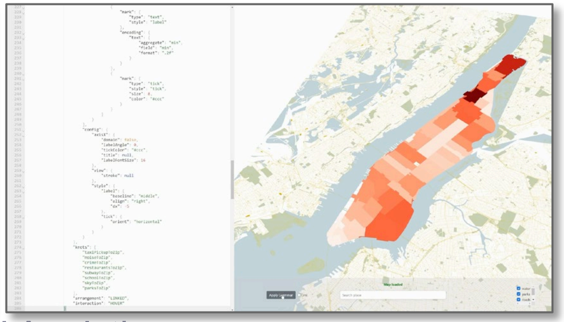
Evaluation

Case studies

- Motivated by real-world problems
- Inspired by previous collaborations

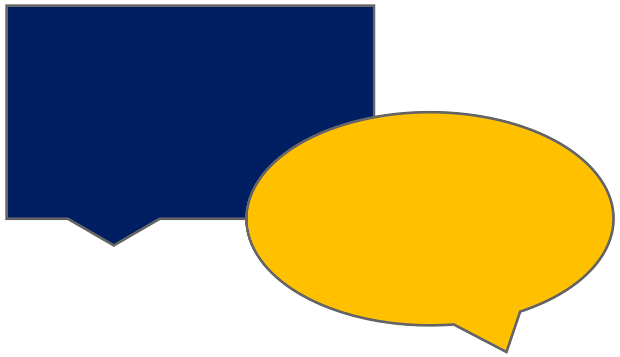




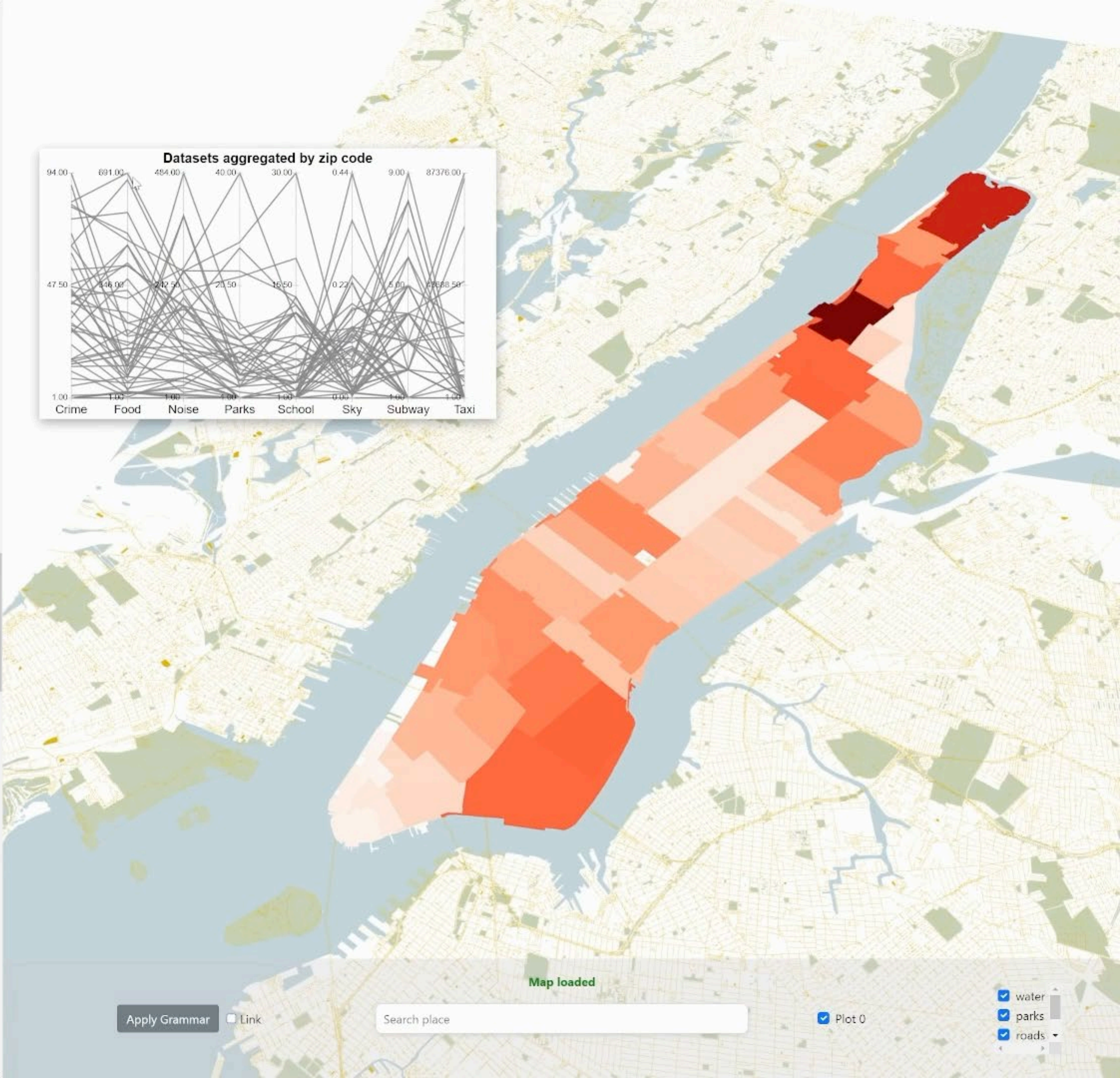


Experts' feedback

- One-hour semi-structured interviews
- Their perspectives on UTK's usability, limitations and needed features



```
                        "aggregate": "min",
                        "field": "min",
                        "format": ".2f"
                    }
                }
            },
            {
                "mark": {
                    "type": "tick",
                    "style": "tick",
                    "size": 8,
                    "color": "#ccc"
                }
            }
        ]
    }
],
"config": {
    "axisX": {
        "domain": false,
        "labelAngle": 0,
        "tickColor": "#ccc",
        "title": null,
        "labelFontSize": 16
    },
    "view": {
        "stroke": null
    },
    "style": {
        "label": {
            "baseline": "middle",
            "align": "right",
            "dx": -5
        },
        "tick": {
            "orient": "horizontal"
        }
    }
}
},
"knots": [
    "taxiPickupToZip",
    "noiseToZip",
    "crimeToZip",
    "restaurantsToZip",
    "subwayToZip",
    "schoolToZip",
    "skyToZip",
    "parksToZip"
],
"arrangement": "LINKED",
"interaction": "HOVER"
}
],
"knots": [
    {
        "id": "noiseToZip",
        "linkingScheme": [
            {
                "predicate": "CONTAINS",
```



text tree table


```json
{
    "views": [
        {
            "map": {
                "camera": {
                    "position": [
                        -9754118,
                        5116849.5,
                        0.6765849609375
                    ],
                    "direction": {
                        "right": [
                            -643.786865234375,
                            -491.20269775390625,
                            676.5849609375
                        ],
                        "lookAt": [
                            449.0103454589844,
                            1677.0311279296875,
                            -1085.390380859375
                        ],
                        "up": [
                            0.26433828473091125,
                            0.524476945400238,
                            0.8093511462211609
                        ]
                    }
                },
                "knots": [
                    "purewater",
                    "pureparks",
                    "pureroads",
                    "opOnSurface",
                    "opOnBuildings"
                ],
                "interactions": [
                    "NONE",
                    "NONE",
                    "NONE",
                    "NONE",
                    "PICKING"
                ]
            },
            "plots": [
                {
                    "plot": {
                        "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
                        "background": "rgb(0,255,0)",
                        "mark": {
                            "type": "arc",
                            "stroke": "black",
                            "strokeWidth": 5
                        },
                        "encoding": {
                            "theta": {
                                "field": "bin",
                                "type": "nominal",
                                "legend": null
```


Experts' feedback

"… facilitates engagement not only across disciplines, but also across urban communities" (Weather scientist)

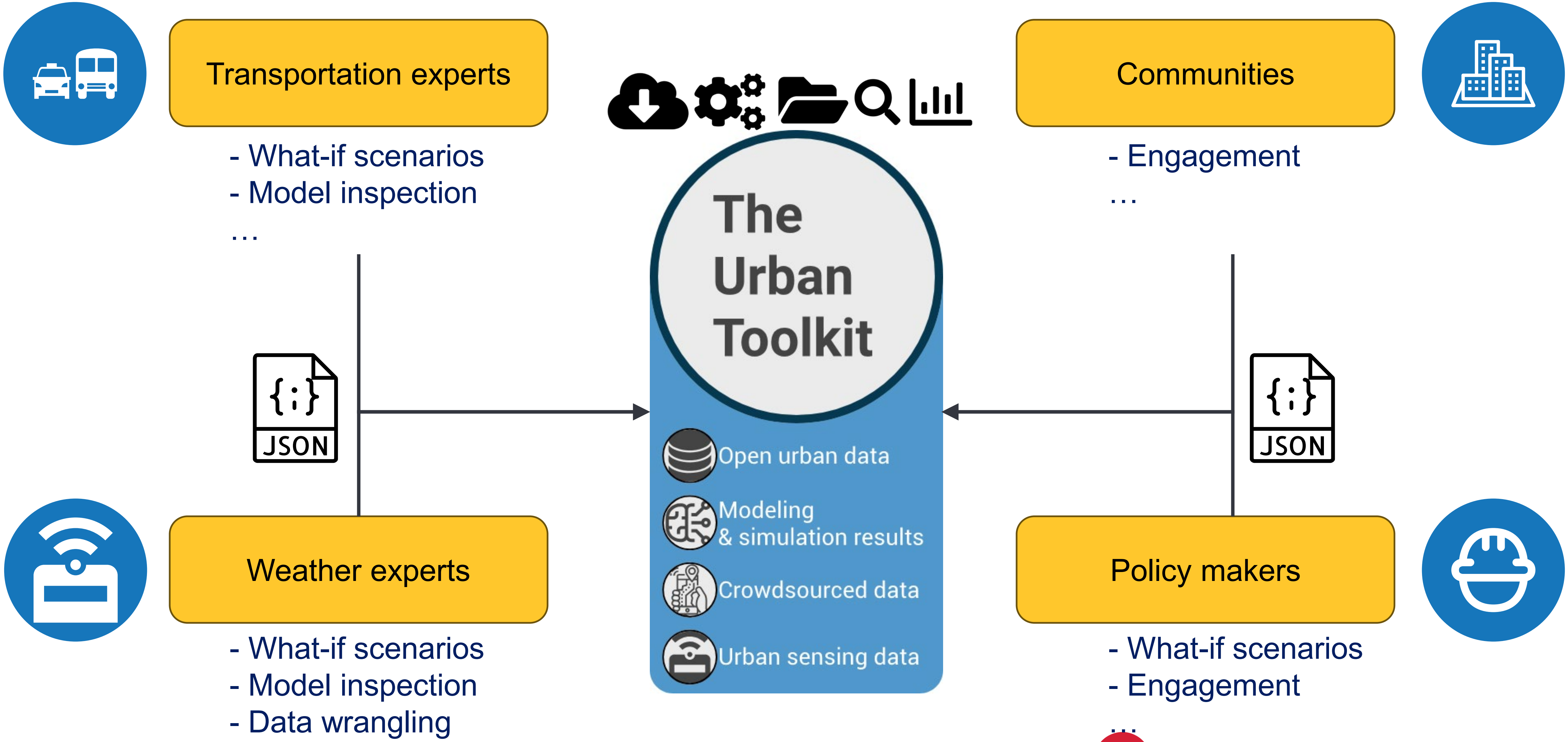
"… researchers could easily share their visualizations instead of cumbersome GIS files" (Urban planner)

"3D makes it more attractive to users and a great tool for communication" (Urban planner)

"It would require training to educate people on the grammar, with examples to showcase its use" (Urban planner)

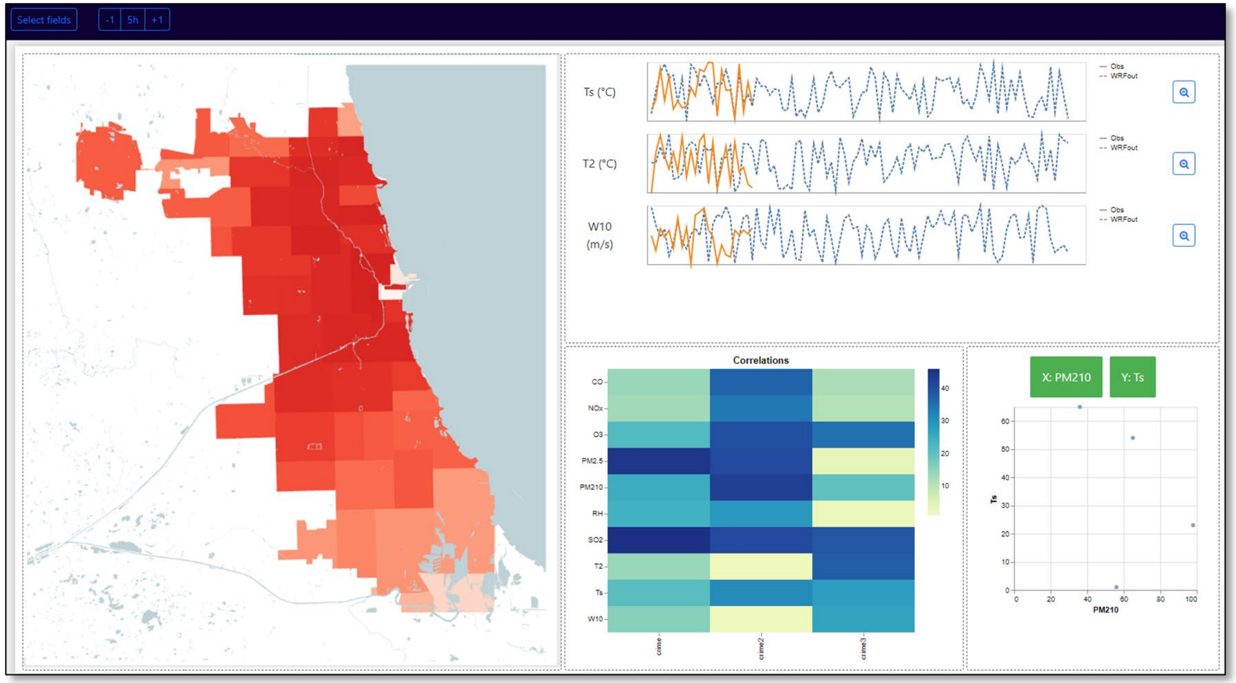
Future opportunities

- Incorporate other functionalities that have been highlighted in previous urban visual analytics works:
    - Computational topology
    - Wavelet
    - Techniques for model inspection


Transportation experts
- What-if scenarios
- Model inspection
…
Communities
- Engagement
…
JSON
JSON
The Urban Toolkit
Open urban data
Modeling & simulation results
Crowdsourced data
Urban sensing data
Weather experts
- What-if scenarios
- Model inspection
- Data wrangling
…
Policy makers
- What-if scenarios
- Engagement
…

Ongoing works extending UTK

Urban accessibility

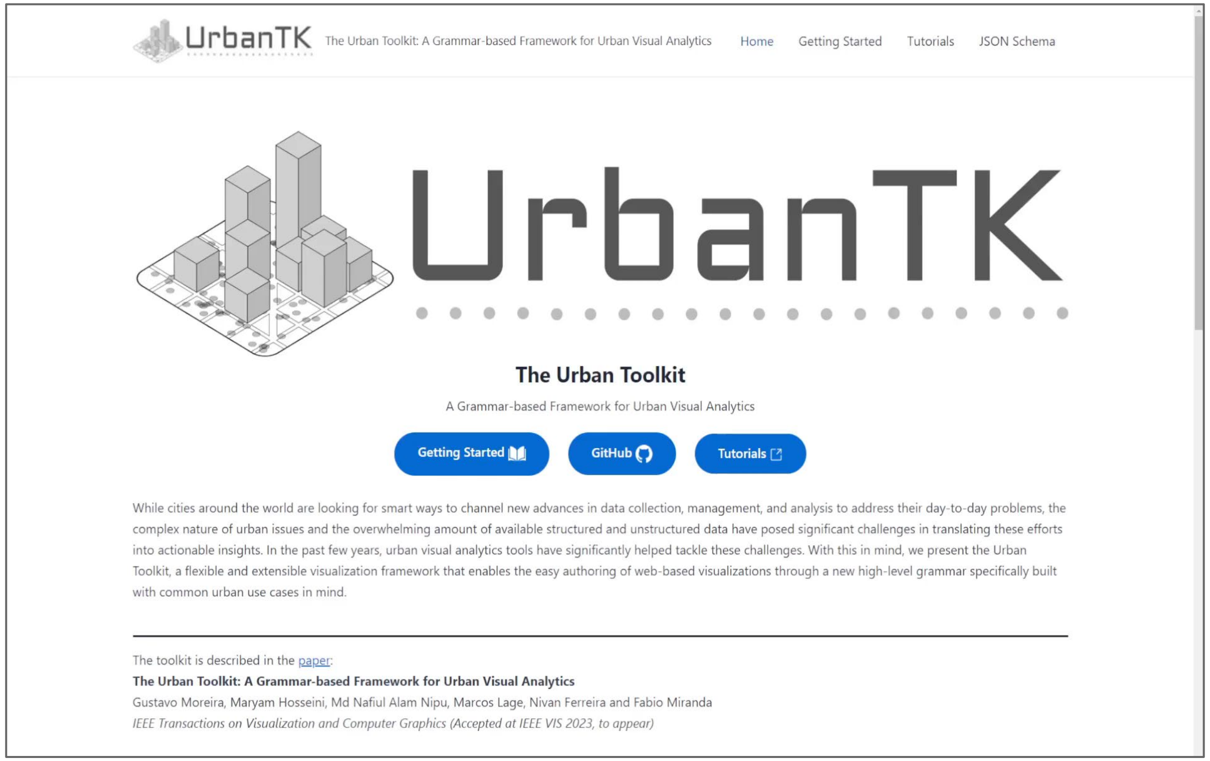


Environmental justice

UrbanTK

Code & tutorials:

urbantk.org



UIC MIT uff VIRTUS IMPAVIDA