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

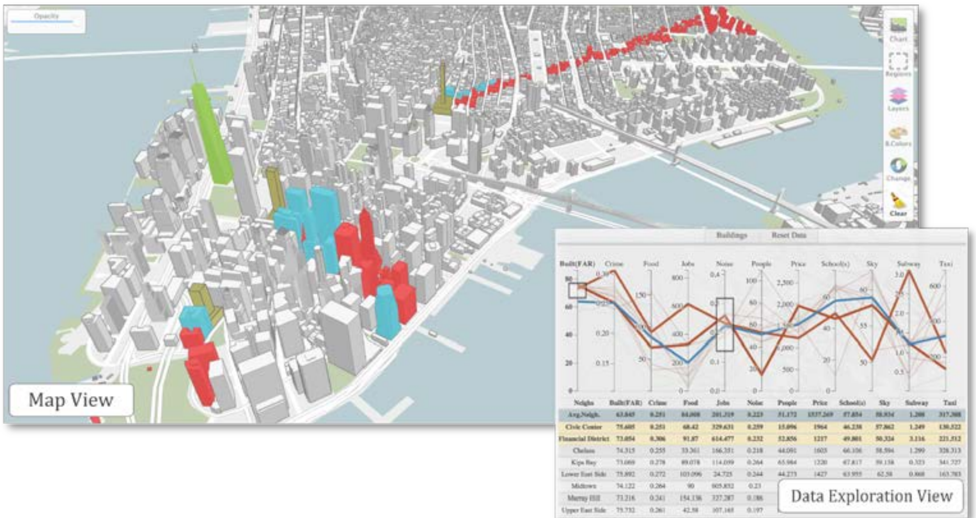

Figure 1: Urbane provides architects, developers, and planners with a new, data and analysis rich way of reading the city with the goal of improving decision making in urban development. Users can explore properties of neighborhoods and buildings using the data exploration view to identify underdeveloped 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

with the impact on the character of a neighborhood. As a result, architects are focused on two issues throughout the design process: a) what defines the character of an existing neighborhood? and b) how will new development change the existing neighborhood?

As more urban data sets become available, opportunities for data-driven approaches to better support the answers to these questions emerge. Through a data-driven understanding of the existing and potential future city, architects, developers, and planners can better collaborate and make more informed decisions. There are several challenges in creating a framework that can facilitate this type of urban decision making.

First, cities are complex environments in which multiple factors play a role in shaping the quality of a particular neighborhood. Therefore, many distinct data sets (of differing dimensionality) need to be considered. Furthermore, such a framework has to support the different stakeholders of development process, such as, architects, developers and planners who often have competing objectives and work at different scales. In this context, it is important to provide flexibility to interactively identify and explore the range of possible developments and measure the effect of these changes. However, assessing the effect of developments requires expensive computation which poses challenges for interactivity.

Tools to perform analysis on urban data sets often visualize the data in two dimensions, however, architects design and communicate to clients and the public in three dimensions. While 3D is often regarded to not be appropriate for visualization [7, 35], many of the properties of interest to the users requires moving away from "flatland", given the 3D data nature of the geometry of buildings and their surroundings. Existing tools used by architects for this purpose are predominantly modeling tools [26] and do not support 2D data sets. Because these tools are designed primarily for modeling, analysis features are typically lacking, and when present, are not integrated with data sets or other analysis features and are time consuming to run. It is therefore important that an analysis framework for architects has the ability to not only support and visualize 2D and 3D data sets in a seamless manner, but also accomplish this efficiently.

Contributions. To address these challenges we propose Urbane, a 3-dimensional framework that enables a data-driven approach for decision making in the design of new urban development. This is the result of a year long collaboration between visualization researchers and architects.

Taking into account the real world requirements of architects, city planners, and developers, we first draft a set of tasks that is to be performed. Urbane was then designed to cater to these analysis tasks. It provides multi-resolution analysis capabilities, i.e., enables experts to analyze the city in different levels of aggregation and thus supports tasks ranging from study of characteristics of large regions (such as neighborhoods) to identifying and simulating opportunities for new developments. In particular, Urbane has the following properties:

- Supports visualization of large collections of both 2D and 3D data sets.
- Support exploration of the data on three different scales – between neighborhoods, within a neighborhood, and with respect to individual building.
- Ability to support "what-if" scenarios and compute impact of the proposed changes in real time.

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. Our collaborators are already using Urbane in their real world projects.

2 Related Work

Understanding cities through data analysis is currently a popular topic with contributions from different communities [2]. Such a wide interest in this area has to do with the importance of cities, the environment inhabited by a majority of the human population. Moreover, the proportion of the world's urban population is expected to grow rapidly in the near future. One of the main challenges in this context is to promote sustainable growth of cities.

In computer graphics, many systems have been developed for modeling and rendering focused on obtaining appealing 3D visuals and simulating attributes of the produced models. One important problem that is investigated in this context is that of urban reconstruction [18]. The research in this area focuses on problems related on recovering buildings geometry from data acquired from sensors [19, 31]. While related, this is orthogonal to our research, since our goal is to perform analysis using models that were already curated. Another related area of research is procedural city design [32], which focuses on parametric techniques to produce complex urban models [23, 33] usually on a city scale. These techniques are used to design entire cities from scratch based on the requirements and focus primarily on the geometric properties of the city.

Multiple visual analytics systems and techniques have been proposed to interactively explore and analyze urban data [11]. It is often the case that these systems are designed to analyze data layers independently. For example, there are individual visual analytics systems in transportation and mobility [1, 8, 36, 38], air pollution [25], real-estate ownership [13, 28] and public utility service problems [39]. Closely related to our paper is the work by Chang et al. [6], in which was designed a tool for exploring multiple urban data sets (at different levels of aggregation). However, while their goals was to explore multiple urban data sets, their system does not support 3D analysis, and hence the ability to evaluate the impact of new developments on their surroundings, which is one of the main requirements of the architects.

Recently, several software platforms have emerged that aim to use urban data sets to help inform the decision making process in the development of cites. First, there are platforms that integrate 2D urban data sets, such as Place I Live [24], aimed at allowing the general public to find neighborhoods and apartments by filtering different urban data sets. Next there are platforms that pair available urban data sets with tools that allow for users to speculate on changes to the city. Transitmix [29] provides a framework for both transit planners and the general public to propose new bus lines, evaluating them based on cost and population served. Finally, there are platforms that allow for procedurally generated 3D buildings, sometimes based on or integrated with urban data sets, in order to test potential development. Flux Metro [9] is a platform that visualizes the development code of Austin, Texas and generates buildings that comply with it. ViziCities [34] allows visualization of procedural 3D models together with other data layers. ArcGIS [16] is a general GIS system which provides 3D analysis capabilities, such as visibility, which are related to our measures of impact. However, it does not scale well with the size of the data being handled, the number of data sets that need to be integrated, and does not provide much capability at the individual building scale making it unsuitable for our purpose.

Our approach to creating an urban analysis framework shares several similarities to these existing platforms: urban data integration, impact analysis, and a 3D visualization environment. In addition, by applying efficient spatial data structures and computer graphics techniques with information visualization techniques, Urbane provides interactive exploration capabilities to understand the urban data at multiple scales as well as perform impact analysis over large collections of data which is not possible in current available systems.

3 Desiderata

In the initial stages of our collaboration, we had several work sessions where we established our objectives and defined the tasks to achieve them. The goal of this work was to design an interactive framework to support the following tasks.

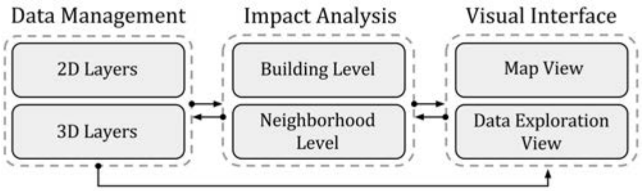

Figure 2: The different components of Urbane and how they interact. The data management component supports the use of both 2D and 3D data layers. The impact analysis component enables the assessment of how new buildings affect their surroundings. The visual interface component supports exploration of the data layers.

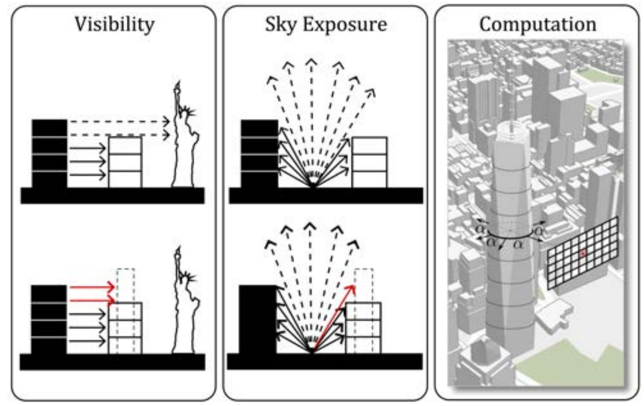

Figure 3: **Visibility.** The white building occludes a portion of the landmark as seen from the black building (top diagram). By changing the white building with the dashed one, the visibility is now totally occluded and the impact is represented by the red rays (bottom diagram). **Sky Exposure.** Dashed rays represent the view directions that can see the sky, while the solid rays represent occluded directions. The resulting impact of changing the white building by the dashed one is represented by the red ray. **Visibility Computation.** The process of computing the visibility to the Freedom Tower (landmark of interest) consists of rendering the scene and marking the buildings that are visible from uniformly distributed view points around that landmark at different heights.

1. *Users should be able to seamlessly explore the city based on multiple 2D and 3D data layers.* Being able to explore multiple 2D urban data layers in the context of actual 3D buildings allows all stakeholders (architects, developers, planners, community board, etc.) to see connections between the design of a building and urban data layers that would not be possible in 2D.
2. *Ability to explore the city at different resolutions, in particular, across neighborhoods and buildings within a neighborhood.* This allows multiple stakeholders to explore different aspects based on their specific objectives. A developer, for example, would want to better understand the characteristics that drive value within a neighborhood to maximize the value of a new building. On the other hand, planners would want to understand differences across neighborhoods and what affects their value in order to plan for future development.
3. *Ability to replace existing buildings in a city with new buildings.* This task allows all stakeholders to evaluate different design options for the a particular project.
4. *Compute the impact of a new building on other buildings.* This task helps users understand the impact of a new building on the surrounding ones. In this scenario, architects are interested in measuring the impact of a new building on the views from existing buildings. In particular, *landmark visibility* from buildings is very important in the development process and is of interest not just to architects (views can inform design), but also to developers (generates value [4]) and city planners (defines neighborhood character).
5. *Compute the impact of a new building on the neighborhood.* This task helps users understand the impact of a new building on the attributes of a neighborhood. In particular architects are interested in measuring the impact of a new building on the *sky exposure* of the surrounding streets. Sky exposure is a critical attribute to measure as it is directly related to available light at the street and thus is closely linked to the perceived quality of a neighborhood.

4 Urbane Framework

We now briefly describe the components of Urbane, shown in Fig. 2, which were designed to support the various tasks of the architects. Urbane consists primarily of three components.

Data Management. Task 1 requires our framework to have the ability to support exploration of different kinds of urban data. We accomplish this through the use of a custom data management component that enables efficient data usage throughout our system. This component supports different types of 2D and 3D data layers that model physical and qualitative aspects of the city. Physical aspects correspond to city infrastructure such as buildings and road-networks, while qualitative ones correspond to measurements associated with quality of life in the city, such as presence of restaurants, noise complaints, and crime, etc. We explain the different data layers supported by our system in detail in Section 5. In order to enable fast retrieval of the data for the rendering as well as computation purposes, we index the data layers using the kd-tree data structure [5].

Impact Analysis. This is the computational component of our framework. Its purpose is to assess the impact of new buildings both on other buildings (Task 4) as well as on the neighborhood (Task 5). To do so, our framework allows users to replace existing building geometries by new ones (Task 3) and compute the change in the sky exposure and landmark visibility caused by this change. In order to perform this computation interactively, we use a rasterization strategy coupled with the use of OpenGL *compute shaders*, which allow general computations as part of the rendering pipeline. The produced impact measures are then mapped to data layers so they can be visualized. Section 6 describes this component in detail.

Visual Interface. This component consists of two main widgets that facilitate visual exploration and analysis of the data layers in our system. The first widget is the *Map View*, which is a 3D map environment that enables the spatial visualization of the data layers loaded in system. In order to achieve interactive rendering, this widget makes use of a number of strategies such as view frustum culling and *tessellation shaders* [17]. The second widget is the *Data Exploration View*, which consists of a parallel coordinates [15] based view whose main purpose is to support visual exploration and filtering across different data attributes. As we explain in Section 7, these widgets allows users to explore the data at multiple resolutions (Task 2).

5 Data Layers

In this section we describe the data layers supported in Urbane and how they can be used to model physical and qualitative aspects of a city. We classify the data layers as 2D and 3D layers, which can be either pre-computed or dynamically derived.

2D Data Layers. We support four types of 2D data layers – point layer, line layer, polygon layer, and grid layer.

Point layer. This layer is used to represent physical aspects of a city associated with locations such as positions of subways stations, as well as, qualitative aspects like noise complaints and crime occurrences.

Line layer. Many physical aspects of a city such as the road or the subway networks are represented as a set of lines. Furthermore, qualitative aspects can be mapped to these lines. For example, as described later in Section 6, the sky exposure measure is computed and visualized along the roads of the city.

Polygon layer. This layer is used to represent regions of interest in a city such as city neighborhoods, parks, and zip codes. It is also possible to associate values to each of these polygonal regions by aggregating point data. For example, the number of crime occurrences can be aggregated for each city neighborhood.

Figure 4: Examples of visualizations of different data layers supported in Urbane. **(a)** Polygon layer showing the different neighborhoods in Manhattan. The polygon corresponding to the Financial District is highlighted (yellow). **(b)** A grid layer representing the density of subway stations in Manhattan. **(c)** Representation of 2D and 3D data layers of the city in a single view. The heat-map along the road network represents the sky exposure along those streets. **(d)** Transparency can be used to avoid occlusion of the 2D layers by the 3D building geometries.

Grid layer. This layer is used to represent data aggregated over a fine grained grid that covers the city. This layer can be seen as a more detailed aggregation compared to polygon layers. For example, noise complaints and crime occurrence densities across the city are examples of useful grid layers.

3D Data Layers. The use of 3D data in the context of a city is critical in the workflow of architects. For example, as mentioned in Section 3, they are not only interested in visualizing buildings' geometries, but are also interested in measuring the impact caused by changes in these geometries on their surroundings. Our framework support two types of 3D layers – parametric meshes and triangle meshes.

Parametric meshes. This type of mesh defines the structure of a building using a grammar, specifying parts of the buildings as set of walls and roofs together with their geometries and textures.

Triangle meshes. They are used to model buildings having a high level of detail such as the different landmarks, and other buildings of interest to the architects which cannot be easily represented using the grammar that is used for parametric meshes.

6 Impact Analysis

One of the main goals of Urbane is to assess the impact of new developments in a city. There are different ways to quantify this impact. In this section, we describe two measures that are commonly used by the architects. We also present strategies to compute these measures interactively and discuss the achieved efficiency.

6.1 Impact Measures

We allow the user to inspect the impact of a new construction with respect to two quantitative criteria, namely, *landmark visibility* and *sky exposure* (illustrated in Fig. 3). These criteria are associated to a particular configuration of buildings in the city. The impact corresponding to each of them is then defined as the difference in the values before and after some buildings are replaced by newer ones.

Landmark visibility. The consideration of views from buildings is important in the development process. While it is hard to quantify the quality of a view, one important aspect usually considered is the visibility of city landmarks [4]. This measure is of relevance to architects since they have to balance the interests of developers, who want to maximize landmark visibility, and city planners who try to minimize the effect of new developments on their surroundings with respect to landmark visibility.

Sky exposure. This criterion measures the percentage of sky that is visible along the streets. While not explicitly measuring direct sunlight, this metric represents access to both daylight and ambient light. City planners want to control sky exposure along streets, which constrains new building projects proposed by developers and architects.

6.2 Computation

In order to compute both landmark visibility and sky exposure we use a rasterization strategy which renders the scene from different points of views and computes these measures based on counts of pixels of the obtained image. The impact is then quantified as the difference in the counts of pixels for the original and changed geometry. We next describe this process in more detail.

Landmark visibility. We can extract the amount of a landmark that can be seen from a particular view point by rendering the scene from that view point and counting the number of pixels in the image corresponding to that landmark. In order to quantify the visibility of a landmark for a large number of buildings, this process would need to be repeated an infeasibly large number of times. To make this computation efficient, we make use of the following key observation – *if a building can view some part of a given landmark, then there exists a view from that portion of the landmark to the building*. The advantage of considering this observation is that, instead of computing views from all the buildings, we only have to compute the views from different points of the landmark. To do so, we render the scene placing the camera at uniform heights over the center of the landmark. At each height, the look-at direction of the camera is set at different angles. This procedure is illustrated in Fig. 3 (Computation). By encoding each building uniquely with a color, it is possible to identify the buildings that are visible in a given view.

Sky exposure. The streets of the city are divided into line segments of equal size, and the sky exposure is computed at the center points of these line segments. It is computed by first placing the camera at each of these points looking upwards. When rendering the scene from the camera, the scene is first cleared so that the entire scene has a unique color corresponding to the sky. Next, all objects / attributes of the scene are colored with a different color. The sky exposure is then computed as the fraction of sky pixels to the total number of pixels.

6.3 Efficiency

The rasterization approach used for computing landmark visibility and sky exposure is composed of two main phases. In the first phase, the scene is rendered from a given camera configuration. This is done efficiently by using the graphics pipeline together with the kd-tree index to support fast clipping queries. The second phase counts the number of pixels in the obtained image that have a certain property (sky color pixels, building pixels). The expensive aspect of this step is in retrieving the rendered scene from the graphics processing unit (GPU) and performing the pixel counting on the CPU, which impacts the interactivity of the application. In order to perform this computation in real time, we make use of the newly introduced compute shader that has been included as part of the graphics pipeline [17]. The compute shader basically allows one to

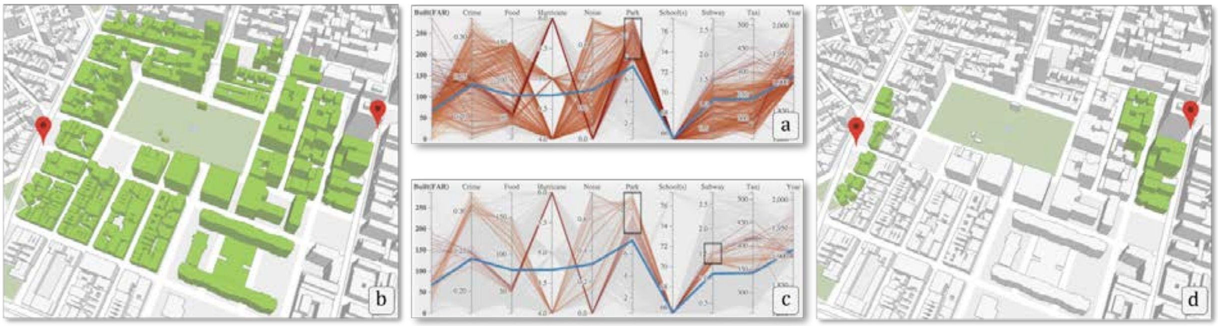

Figure 5: Exploring the city at multiple scales using the data exploration view. **(a)** The user first selects only buildings close to a park using the PCC (the value corresponds to the area of park space weighted by distance). **(b)** The buildings satisfying the constraints are highlighted in the map view. **(c)** The user now selects those sites with high density of subway near them. **(d)** The buildings remaining after this filter is applied.

perform non-graphics (or GPGPU) operations while still being part of the rendering pipeline. Since it has access to all the buffers and textures used by the vertex and fragment shaders, there is no need to transfer the data between the CPU and GPU. Further, the required count operations are easily parallelizable. Therefore these can be efficiently accomplished by using the hundreds of cores available on modern GPUs.

In order to have an idea of the speed-up obtained by the use of compute shaders, we compare the performance of using a CPU with that of a GPU in the second phase. In our experiments, the scene is rendered to a 256x256 image. Using this setup, computing sky exposure at 650 locations takes 8.4 s using the CPU. Using the compute shader, we accomplish the same task in 75 ms, providing two orders of magnitude (112x) speed-up. All experiments were run on a desktop with an Intel Xeon E5-2650 CPU, 32 GB RAM, and a Nvidia GTX 680 graphics card.

7 Visual Exploration Interface

We worked closely with the architects in the design of Urbane's user interface in order to support the tasks described in Section 3 and provide an intuitive user experience. The visual interface of Urbane is composed of two components, *Map view* and *Data Exploration view*, illustrated in Fig. 1.

Map view. This view is composed of a 3D map rendering component. Overlaid menus and panels are used in order to maximize the map rendering area of the screen real estate. We support two possible states of map rendering – 2D and 3D. In the 2D state (Figs. 4 (a) and 4 (b)), a top view of the map is shown similar to conventional GIS map interfaces. This state is used to visualize the 2D data layers. The 3D state visualizes both 2D and 3D data layers. For example, in Fig. 4 (c) the 2D layers representing physical aspects of the city are shown together with a heat-map denoting the sky exposure over the road network and a 3D layer representing the geometry of the buildings. As shown in Fig. 4 (d), transparency on the 3D layers can be used to avoid occlusion. The level of transparency can be adjusted by using the opacity slider (top left of map view in Fig. 1). Navigation and operations on Map view such as panning, zooming, and rotating the view are accomplished through mouse interactions. The main menu (right side of map view in Fig. 1) allows users to control all the functionalities of the system including that of activating the different data layers, performing impact analysis, and toggling the Data Exploration View.

Data Exploration View. The main goal of Data Exploration View is to support the analyses of urban data representing qualitative aspects of the city in two resolutions – at *neighborhood* and *building* levels. This view is composed of two components – a parallel coordinates chart (PCC) and a data table. While the PCC allows users to analyze and compare multiple entities (neighborhoods or buildings) with respect to each other, the data table helps them view the precise values corresponding to entities of interest. The value for each building is computed by using the weighted sum approach described in Section. 8.1. The records corresponding to the different neighborhoods are obtained by computing the average value for each data dimension over all buildings in the neighborhood. Urban planners and developers use analyses at the neighborhood resolution to understand the characteristics of both a single neighborhood, as well as differences between neighborhoods. Once they decide on neighborhood(s) of interest, they can then perform the analysis at the resolution of buildings. This is done by selecting the *Buildings* option on the top of the widget.

Each qualitative 2D data layer corresponds to one dimension in the PCC. Users can interactively toggle on or off data layers of interest. Users can also modify the properties of the PCC such as reorder the dimensions (to explore correlations among the different dimensions), color code different lines based on a data set, and flip range of the axes. Each line that is visualized corresponds to either a single neighborhood or building, depending on the resolution. In addition, for comparison purposes, we also visualize the attributes corresponding to the average of the items being shown (which is highlighted in blue in the PCC). The PCC in Fig. 1 visualizes the data at the neighborhood resolution, while the ones in Fig. 5 visualizes the data at a building resolution.

Interacting with the data. The main exploration workflow supported in Urbane consists first in exploring the urban data at the neighborhood level and later drilling-down to the building level to identify possible development locations. In order to do so, the Data Exploration View can be used to select and filter entities having the required range of values along different data sets. The filtered entities are listed in the data table and are also highlighted on the map view. Either the selected neighborhoods (Fig. 4 (a)) or the selected buildings are highlighted depending on the resolution. Consider the example in Fig. 5. Here the user first filters buildings that are distant from parks (Fig. 5(a)). The result of this filtering is shown in Fig. 5(b), where the buildings close to a park are highlighted. Additional filtering to remove regions having a lower density of subways (Fig. 5(c)), results in selecting only buildings close to the two subway stations highlighted in Fig. 5(d).

Testing new developments. Once a building of interest is chosen, the user can replace it with a new mesh using the *Change* button (Fig. 1). Users can pre-load a set of pre-defined meshes among which one is chosen as a replacement. This operation will trigger the impact analysis computation. The resulting impact on landmark visibility is shown by appropriately coloring the affected buildings, as shown in Fig. 1. Buildings which have the landmark visibility decreased are colored red, while buildings for which the landmark visibility improves are colored blue. Similarly, the impact in sky exposure is shown by coloring the affected portions of the streets as shown in Fig. 7.

8 Use Case Scenarios

To demonstrate the capacity of Urbane we present a use case in which we, the architects, assist a developer in identifying a site in New York City as well as evaluate different buildings designs according to the impact measures previously defined. We start by describing the data sets used in this use case.

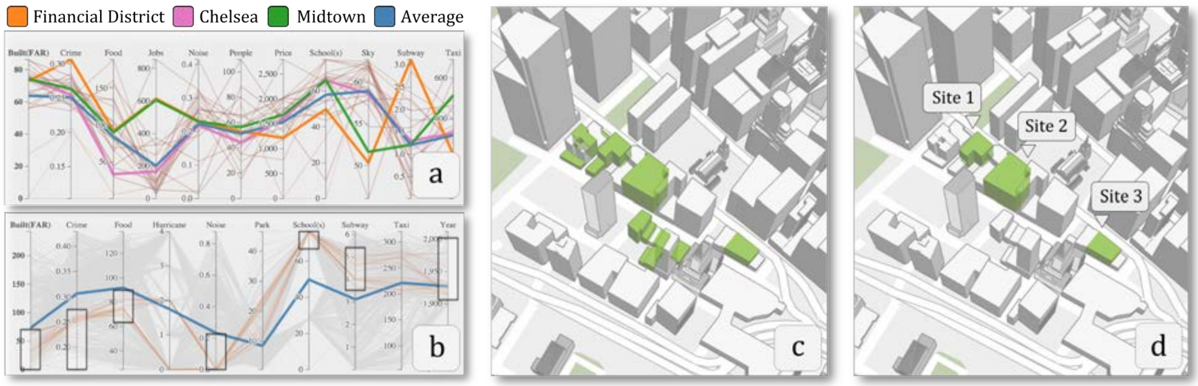

Figure 6: Using Urbane to identify development sites in Financial District. The data exploration view of Urbane is used to study the characteristics of Financial District with respect to other neighborhoods in Manhattan (a). This is then used to filter (b) and identify potential development sites (c). Further filtering based on the site properties isolates three sites (d) that have high development potential.

8.1 Data Setup

We used a diverse collection of urban data sets from New York City that support decision making in the design and development process. In a pre-processing step, these data sets are converted into a set of layers that can be loaded into Urbane.

Physical data layers. For the physical aspects of the city, such as the geometry of land, streets, parks, water bodies we currently use data from Open Street Maps [37]. For important buildings, such as landmarks, we generated and used high resolution meshes, represented as triangle meshes. For the rest of the buildings, we use the parametric meshes also obtained from Open Street Map.

Qualitative data layers. Data sets describing qualitative aspects of Manhattan span all the data layers types supported in Urbane.

Point data. Data corresponding to locations of crime occurrences, taxi activity, subway stations, noise complaints, and restaurants (obtained from [20, 21]) are available as point data.

Line data. The sky exposure along the streets of Manhattan is precomputed and represented as a line layer. It is computed at 10-meters intervals along all the streets of the city.

Polygon data. Population density, jobs density, building density [20], and average price of properties [30] are available as values for each neighborhood in New York City. Other polygonal data used in our analysis are hurricane evacuation zones [21], parks, and elementary school zones [21] with the corresponding school quality report [20].

Grid data. Point data was used to derive grid layers as follows. Manhattan was first partitioned into a grid of square cells having width 164 ft (50 meters). Then, given a data set, for each cell, we add up the values obtained by applying a Gaussian Kernel to the cell center and each point within a radius of 0.25 miles (5 min walking distance) from the cell center. Intuitively, this counts the occurrences of the entity in the point data, and thus provides a proximity function for that entity.

Impact layers. The impact of landmark visibility (Fig. 1) and sky exposure (Figs. 4 (c) and 4 (d)) are computed in real time when the user changes the configuration of the city.

8.2 Use Case Overview

In this use case we focus on the Financial District neighborhood (highlighted in Fig. 4 (a)) to identify and develop a residential building. It is one of the oldest neighborhoods in New York, is extremely dense and has an irregular street grid creating many unique and difficult to develop sites. Given this complexity an architect will need strong understanding of the neighborhood characteristics to help identify sites for development and eventually, facilitate the negotiation process with the city planner, who cares about maintaining the quality of the neighborhood. The developer wants to maximize the value of a development while a city planner wants to mitigate the negative impact of new developments. The architect must reconcile these competing objectives.

8.3 City Scale: Understanding Financial District

First, we use Urbane to understand the Financial District neighborhood in the broader context of other neighborhoods in Manhattan. By comparing with other neighborhoods we can understand its strengths and weaknesses and establish performance thresholds from other well-known and well performing neighborhoods.

The attributes of Financial District, the orange line in Fig. 6(a), are surprisingly close to Manhattan averages with a few exceptions. The values for job density and subway access (see Fig. 4 (b)) are better than the average, while sky exposure is much lower. This illustrates strengths in job and transit access and a need to be sensitive by not reducing sky exposure with new development. Note that neighborhoods having high crime typically have a low job density. However, there is high crime in Financial District even though it has a high job density, perhaps indicating that a lack of 24 hour activity (all jobs and little residential) is linked to crime. Midtown, the other primary business district, is the neighborhood having the most similar characteristics across data sets (green line in Fig. 6(a)).

While no neighborhood performs better than the Manhattan average on all the attributes, Chelsea has the best overall performance (pink colored line in Fig. 6(a)). This supports our expectations as Chelsea is generally understood as a desirable and quintessential New York neighborhood. We use it as reference for neighborhood performance.

The above analysis of neighborhood characteristics suggests that when looking for sites to develop in Financial District, transit access is not an issue, but crime is, and any new development needs to be sensitive to sky exposure impact.

8.4 Neighborhood Scale: Filtering Sites

The next step in our process is to use the understanding of Financial District attributes and how they relate to other neighborhoods in Manhattan to identify sites that have development potential. In order to identify such sites, we developed the following criteria for filtering different attributes.

For all attributes, except for Built (FAR) and Year (which refers to year of construction), we filter for sites that are better than the neighborhood average. Built (FAR) is the percentage of the maximum allowed area for the site that is actually built. We select sites with Built (FAR) 55% of allowed capacity or less because it is unlikely that a building having a higher capacity will be torn down for new development. The Built (FAR) range exceeds 100% because many of the buildings in the Financial district were built before area was regulated. We filter Year to include only buildings older than 2000 as newer buildings are unlikely to be redeveloped. We do not consider food and parks attributes (i.e., density of restaurants and parks respectively) in the filtering step because a new building can address these by adding a grocery store or public space. All of the applied filters are illustrated in Fig. 6(b). This results in 15 potential

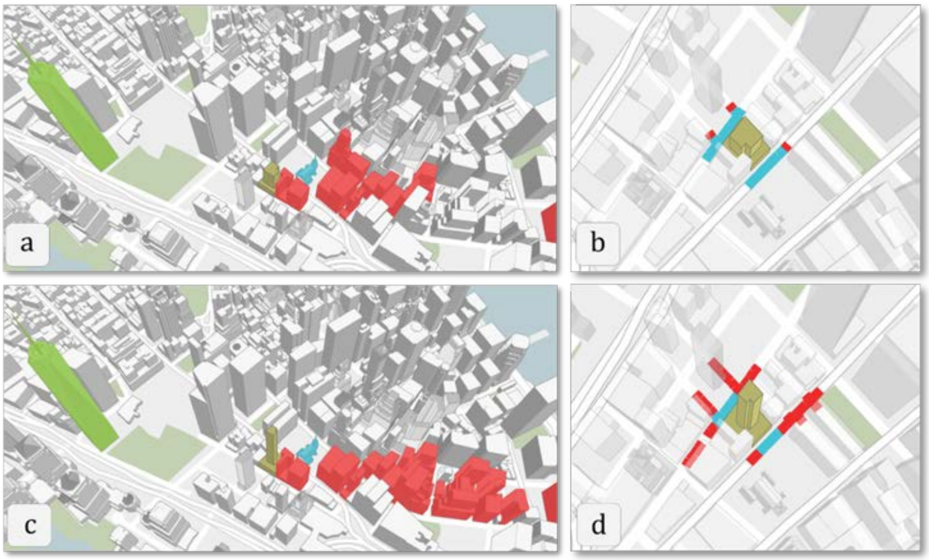

Figure 7: Understanding the impact of different building designs. The view (a) and sky exposure (b) impact when using a 80×120 floor plate vs. the impact when using a 65×65 floor plate (c & d) for the proposed buildings.

#	Area (sq ft)	Floor Plate	Neighborhood Impact			
			% Sky Exposure	Landmark Visibility		
				%	+ve	-ve
1	109,890	80×120	-0.56	-0.38	1	8
		65×65	-0.65	-0.36	1	9
2	268,000	80×120	+0.52	-0.19	1	15
		65×65	+0.12	-0.47	1	30
3	114,700	80×120	-0.75	-0.32	0	6
		65×65	-0.87	-0.27	0	6

Table 1: Results of the analysis from the three identified sites. A positive impact value implies that the building improves the neighborhood. The +ve column of view denotes the number of buildings for which the view has increased, while the -ve column denotes the number of buildings for which the view decreased. The change in sky exposure is computed as the average relative change in the sky exposure measure in the neighborhood. The change in landmark visibility is computed as the average relative change in the view (in terms of pixel count) to the landmark over all impacted buildings.

sites for development primarily concentrated south of the Freedom Tower. These locations are shown in Fig. 6(c).

8.5 Building Scale: Testing Development

For the fifteen sites identified it is important to understand the trade-offs between the value of potential development and the impact on the surrounding context. Twelve of the sites are too small to fully utilize the maximum allowed area or to have commercially viable floor plate size and can be eliminated from consideration. We next use Urbane to study the potential development of the three remaining sites, shown in Fig. 6(d), as residential buildings with ground floor retail. For each of these sites, we load meshes with two different tower floor plate sizes – a 65ft×65ft floor plate for a slender tower and a more typical 80ft×120ft floor plate.

Table 1 summarizes the results of the impact analysis and illustrates the trade-offs between the sites. Site 1 has a modest impact on the neighborhood, while Site 3 performs poorly across all measures and can be eliminated from consideration. Both building scenarios on Site 2 are promising with high performance relative to different attributes. While the 80ft×120ft option impacts the view of 15 buildings, the average view percentage impacted is the smallest (Fig. 7(a)). This might seem counter intuitive since this option has the largest floor area among all the sites, and thus demonstrates the utility of Urbane. The 65ft×65ft option has a low impact on sky exposure (Fig. 7(d)), but the worst impact on other buildings' views (Fig. 7(c)).

The impact analysis therefore reveals that no single site is clearly the best, but that each has strengths and weakness that must be considered. Understanding the trades-offs between various options will allow us (architects) to explain and reconcile the objectives of the developer and city planner. As demonstrated in this use case, using real constraints and a real world context, Urbane provides an effective visual analytic platform for stakeholders of a project to understand the trade-offs between various development scenarios. This results in a better development process as each side can make more informed and defensible decisions rather than arguing a position without knowing the true impact of a development.

9 Expert Feedback

Architects using Urbane have identified several benefits. First, the visualization and responsiveness of the interface allows the architect to pose and test many different questions quickly. Normally this type of analysis requires the use of many tools over several days, rather than in a single tool and in a matter of minutes as with Urbane. The integration of data sets, analysis and multiple scales allows for insights on complex problems that otherwise would not have been possible. Architects pointed out that Urbane allowed them to easily establish assumptions and select criteria to test quickly and get meaningful results. In particular, they found the use of parallel coordinates extremely intuitive and powerful for their needs, which is reflected in the following comment: *Filtering of the parallel coordinates chart allows for easy identification of direct and indirect correlations between data sets that can provide insights into neighborhood characteristics that is not possible with other softwares. Additionally, the ability to visualize the neighborhood and buildings being filtered in real time makes a connection between the data and the city that can better inform the filtering and analysis process.*

The value of Urbane to architects is reflected in the following comment made when working on the use case in Section 8: *"Urbane is extremely fast considering it is displaying and analyzing large 2D and 3D urban data sets, allowing us to test many scenarios seamlessly in multiple scales. The clear and intuitive interface had us engaged very quickly and we believe that anybody, even those without design experience or fluency with analysis programs, can easily use and benefit from Urbane. We think that these aspects make the core concept of Urbane tangible and realistic"*.

As an added attribute to consider the architects suggested including residential sale value by unit to provide a more nuanced understanding between the other attributes and value. As a tool for multiple stakeholders, they suggested that we should provide for a way in which Urbane can also be tested by developers, city planners, and the general public interested in future development. This will help to add or adjust functionality from their point of view.

10 Discussion and Future Work

Extension to other cities. The Urbane framework is general and can be used to perform analysis in any city. In order to do so, data sets representing the physical and the qualitative aspects of the city of interest need to be obtained. For the physical aspects, such as the land, streets, parks, and water bodies we used data from Open Street Maps [37], which includes all the major cities in the world.

In case of the qualitative aspects, many cities are now making available open data gathered from various city agencies [10, 12, 14, 22, 27]. Most of the data sets are provided in a tabular format [3] that can be easily be converted into Urbane data layers.

Impact analysis. In addition to spatial impact measures, it would be interesting to include qualitative attributes in the evaluation of the impact of a proposed change. For example, new office buildings will generate both job opportunities and need for better transportation in the area. This will help enable a data-driven prediction of these attributes and their impact on the value of the neighborhood. Such an analysis can also help planners to identify the kind of buildings that can change certain attributes of a neighborhood.

Evaluation. In this work we designed the interface and visualizations present in Urbane such that it could be easily understood by the architects. Prior to making Urbane public, we intend to do a rigorous user study involving all the stakeholders (*i.e.* architects, developers, and planners) to evaluate how Urbane performs for their varying tasks. We are also exploring different visualizations and how they would support the different tasks along the lines of Dubel et al. [7].

Use of temporal data. While many data sets have a temporal component, we make use of just the spatial information while visualizing the data. In future, we plan to extend our framework to support both interactive querying and visualization of spatio-temporal data.

New building design. For future development, architects would like to have the ability to automatically generate new buildings following the zoning regulations of the city. They also suggested that being able to optimize such building forms relative to specified attributes would greatly help influence the actual design process.

Acknowledgements

This work was supported in part by a Google Faculty Award, an IBM Faculty Award, the Moore-Sloan Data Science Environment at NYU, the NYU School of Engineering, the NYU Center for Urban Science and Progress, Kohn Pedersen Fox Associates, AT&T, NSF award CNS-1229185, and the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

[1] G. Andrienko and N. Andrienko. Spatio-temporal Aggregation for Visual Analysis of Movements. In *IEEE VAST*, pages 51–58. IEEE, 2008.

[2] S. Arietta, A. Efros, R. Ramamoorthi, and M. Agrawala. City Forensics: Using Visual Elements to Predict Non-Visual City Attributes. *IEEE TVCG*, 20(12):2624–2633, Dec 2014.

[3] L. Barbosa, K. Pham, C. Silva, M. Vieira, and J. Freire. Structured Open Urban Data: Understanding the Landscape. *Big Data*, 2(3), 2014.

[4] E. Benson, J. Hansen, J. Schwartz, ArthurL., and G. Smersh. Pricing Residential Amenities: The Value of a View. *The Journal of Real Estate Finance and Economics*, 16(1):55–73, 1998.

[5] J. L. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM*, 18(9):509–517, 1975.

[6] R. Chang, G. Wessel, R. Kosara, E. Sauda, and W. Ribarsky. Legible Cities: Focus-Dependent Multi-resolution Visualization of Urban Relationships. *IEEE TVCG*, 13(6):1169–1175, Nov 2007.

[7] S. Dübel, M. Röhlig, H. Schumann, and M. Trapp. 2D and 3D Presentation of Spatial Data: A Systematic Review. In *Proc. of the IEEE VIS 2014 International Workshop on 3DVis: Does 3D really make sense for Data Visualization?*, 2014.

[8] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual Exploration of Big Spatio-temporal Urban Data: A Study of New York City Taxi Trips. *IEEE TVCG*, 19(12):2149–2158, 2013.

[9] Flux Metro. `https://flux.io/metro/`.

[10] Freedom of Information Act (FOIA), 2014. http://www.foia.gov.

[11] T. Glander and J. Dllner. Abstract Representations for Interactive Visualization of Virtual 3D City Models. *Comput. Environ. Urban Syst.*, 33(5):375 – 387, 2009.

[12] B. Goldstein and L. Dyson. *Beyond Transparency: Open Data and the Future of Civic Innovation*. Code for America Press, 2013.

[13] T.-A. Hoang-Vu, V. Been, I. G. Ellen, M. Weselcouch, and J. Freire. Towards Understanding Real-Estate Ownership in New York City: Opportunities and Challenges. In *Proc. DSMM*, pages 1–2. ACM, 2014.

[14] J. Höchtl and P. Reichstädter. Linked Open Data - A Means for Public Sector Information Management. In *Electronic Government and the Information Systems Perspective*, volume 6866 of *Lecture Notes in Computer Science*, pages 330–343. Springer, Berlin Heidelberg, 2011.

[15] A. Inselberg and B. Dimsdale. Parallel Coordinates. In *Human-Machine Interactive Systems*, pages 199–233. Springer, 1991.

[16] K. Johnston, J. M. Ver Hoef, K. Krivoruchko, and N. Lucas. *Using ArcGIS Geostatistical Analyst*, volume 300. Esri Redlands, 2001.

[17] J. Kessenich, D. Baldwin, and R. Rost. The OpenGL Shading Language (version 4.40). Technical report, 2014.

[18] P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. Gool, and W. Purgathofer. A Survey of Urban Reconstruction. *CGF*, 32(6):146–177, 2013.

[19] L. Nan, A. Sharf, H. Zhang, D. Cohen-Or, and B. Chen. SmartBoxes for Interactive Urban Reconstruction. *ACM TOG*, 29(4):93, 2010.

[20] NYC Department of Planning. `http://www.nyc.gov`.

[21] NYC Open Data. `https://nycopendata.socrata.com/`.

[22] Chicago Data Portal. `https://data.cityofchicago.org`.

[23] Y. I. Parish and P. Müller. Procedural Modeling of Cities. In *Proc. SIGGRAPH*, pages 301–308. ACM, 2001.

[24] Place I Live. `http://placeilive.com/`.

[25] H. Qu, W.-Y. Chan, A. Xu, K.-L. Chung, K.-H. Lau, and P. Guo. Visual Analysis of the Air Pollution Problem in Hong Kong. *IEEE TVCG*, 13(6):1408–1415, 2007.

[26] Rhinoceros. `https://www.rhino3d.com/`.

[27] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, H. Wendy, and M. Schraefel. Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, 27(3):16–24, 2012.

[28] G. Sun, R. Liang, F. Wu, and H. Qu. A Web-based Visual Analytics System for Real Estate Data. *Sci. China Info. Sci.*, 56(5):1–13, 2013.

[29] Transitmix. `http://www.transitmix.net/`.

[30] Trulia. `http://www.trulia.com/`.

[31] C. A. Vanegas, D. G. Aliaga, and B. Benes. Automatic Extraction of Manhattan-World Building Masses from 3D Laser Range Scans. *IEEE TVCG*, 18(10):1627–1637, Oct 2012.

[32] C. A. Vanegas, D. G. Aliaga, P. Wonka, P. Müller, P. Waddell, and B. Watson. Modelling the Appearance and Behaviour of Urban Spaces. *CGF*, 29(1):25–42, 2010.

[33] C. A. Vanegas, I. Garcia-Dorado, D. G. Aliaga, B. Benes, and P. Waddell. Inverse Design of Urban Procedural Models. *ACM TOG*, 31(6):168, 2012.

[34] ViziCities. `http://vizicities.com/`.

[35] B. Wang and K. Mueller. Does 3D Really Make Sense for Visual Cluster Analysis? Yes! In *Proc. of the IEEE VIS 2014 International Workshop on 3DVis: Does 3D really make sense for Data Visualization?*, 2014.

[36] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering. Visual Traffic Jam Analysis Based on Trajectory Data. *IEEE TVCG*, 19(12):2159–2168, 2013.

[37] O. Wiki. Using OpenStreetMap — OpenStreetMap Wiki, 2015.

[38] W. Zeng, C.-W. Fu, S. Arisona, A. Erath, and H. Qu. Visualizing Mobility of Public Transportation System. *IEEE TVCG*, 20(12):1833–1842, 2014.

[39] J. Zhang, E. Yanli, J. Ma, Y. Zhao, B. Xu, L. Sun, J. Chen, and X. Yuan. Visual Analysis of Public Utility Service Problems in a Metropolis. *IEEE TVCG*, 20(12):1843–1852, Dec 2014.