Data-Driven Park Planning: Comparative Study of Survey with Social Media Data

Jisoo Sim

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Architecture Design Research

Patrick Miller, Co-Chair
Cermetrius Lynell Bohannon, Co-Chair
Thomas Skuzinski
Samarth Swarup
Robert Schubert

November, 2019

Blacksburg, Virginia

Abstract

The purpose of this study was (1) to identify visitors' behaviors in and perceptions of linear parks, (2) to identify social media users' behaviors in and perceptions of linear parks, and (3) to compare small data with big data. This chapter discusses the main findings and their implications for practitioners such as landscape architects and urban planners. It has three sections. The first addresses the main findings in the order of the research questions at the center of the study. The second describes implications and recommendations for practitioners. The final section discusses the limitations of the study and suggests directions for future work.

This study compares two methods of data collection, focused on activities and benefits. The survey asked respondents to check all the activities they did in the park. Social media users' activities were detected by term frequency in social media data. Both results ordered the activities similarly. For example social interaction and art viewing were most popular on the High Line, then the 606, then the High Bridge according to both methods. Both methods also reported that High Line visitors engaged in viewing from overlooks the most. As for benefits, according to both methods vistors to the 606 were more satisfied than High Line visitors with the parks' social and natural benefits. These results suggest social media analytics can replace surveys when the textual information is sufficient for analysis.

Social media analytics also differ from surveys in accuracy of results. For example, social media revealed that 606 users were interested in events and worried about housing prices and crimes, but the pre-designed survey could not capture those facts. Social media analytics can also catch hidden and more general information: through cluster analysis, we found possible reasons for the High Line's success in the arts and in the New York City itself. These results involve general information that would be hard to identify through a survey.

On the other hand, surveys provide specific information and can describe visitors' demographics, motivations, travel information, and specific benefits. For example, 606 users tend to be young, high-income, well educated, white, and female. These data cannot be collected through social media.

Keywords: linear park, survey, small data, social media, big data analytics

Abstract for a general audience

Turning unused infrastructure into green infrastructure, such as linear parks, is not a new approach to managing brownfields. In the last few decades, changes in the industrial structure and the development of transportation have had a profound effect on urban spatial structure. As the need for infrastructure, which played an important role in the development of past industry, has decreased, many industrial sites, power plants, and military bases have become unused. This study identifies new ways of collecting information about a new type of park, linear parks, using a new method, social media analytics. The results are then compared with survey results to establish the credibility of social media analytics. Lastly, shortcomings of social media analytics are identified. This study is meaningful in helping us understand the users of new types of parks and suggesting design and planning strategies. Regarding methodology, this study also involves evaluating the use of social media analytics and its advantages, disadvantages, and reliability.

Table of Contents

Tables

Figures

CHAPTER I. INTRODUCTION

1. Background

Big data[1] is emerging as a potentially valuable source of data and may come to play an important role in urban interpretation and planning. For example, sensing urban lands using social media provides the characteristics of urban use, urban area, and urban imaginaries (Frias-Martinez & Frias-Martinez, 2014; Kelley, 2013; R. Lee, Wakamiya, & Sumiya, 2013). Social media data in particular has the potential to improve our understanding of human needs and desires (Afzalan & Muller, 2014; Georgiou, 2017) and their attitudes toward a place and its attributes (Hollander & Shen, 2017). Because users post their activities and feelings on social media to communicate (Kankanhalli, Tan, & Wei, 2005), social media has the potential to be used in design and planning as a tool for understanding the needs of the people who use the landscape. Social media data also don't have geographical limitations, which means researchers can collect data from users worldwide at their desks. Finally, social media data allows researchers to be predict future trends from past datasets.

Turning unused infrastructure into green infrastructure, such as linear parks, is not a new approach to managing brownfields. In the last few decades, changes in the industrial structure and the development of transportation have had a profound effect on urban spatial structure. As the need for infrastructure, which played an important role in the development of past industry, has decreased, many industrial sites, power plants, and military bases have become unused. The same is true of transportation infrastructure. These unused sites are called brownfields, and using

[1] Provost and Fawcett (2013) define big data as data sets large for traditional data-processing systems and that, therefore, require new technologies. According to Zikopoulos, deRoos, Parasuraman, Deutsch, Giles and Corrigan (2013), big data have four dimensions: (1) volume, (2) velocity, (3) variety, and (4) veracity (Zikopoulos et al., 2013).

them as parks is one strategies for making brownfields reusable. Various types of brownfields, such as factories, power plants, military bases, and closed tracks, are turned into parks. There are several reasons for this trend. The first is the goal of restoring brownfields, which often have contamination that prevents uses like housing. The second is expected economic benefits. The third is to provide green space to the public.

Replacing unused infrastructure with parks contributes to the economic growth of communities and cities. This type of urban redevelopment has two benefits for adjacent areas: (1) economic stimulus (the redevelopment of brownfields), and (2) the creation of green space (De Sousa, 2006). Unused transportation infrastructure has been viewed as a drawback for adjacent communities, potentially exposing people to health and safety risks (Ferretti & Degioanni, 2017). However, recent studies of the impact of brownfield redevelopment have claimed economic returns related to taxes, property, and employment (De Sousa, 2006; Simons & Jaouhari, 2001). Reusing unused infrastructure as green space also benefits surrounding neighborhoods by providing places for recreation and relaxation (Crompton, 2001). Many cities propose redevelopment projects that replace unused sites with green space for economic growth.

For these reasons, this study focuses on these two issues: (1) the emergence of big data, especially social media data, and its ability to reveal hidden relationships between environments and behaviors in landscape architecture, and (2) the increasing redevelopment from brownfields to parks, especially linear parks. by comparing big data analytics and small data analytics. Through social media analytics, we try to understand visitors' behaviors and their attitudes toward design, and to plan a new type of urban park, a linear park.

2. Statement of Problem and Research Purpose

A lack of information on linear parks could cause the design and planning process to fail to serve the interests of potential visitors. For a long time, the physical features of parks, such as size, facilities, and vegetation, have been decided by landscape designers according to social context. At the beginning of the nineteenth century, parks were pleasure gardens that brought natural landscape features into the city (Hayward, 1989). Later, physical activities became their main feature (Cranz, 1978), with the purpose of reducing health problems such as obesity and diabetes. To create a park that meets users' needs, designers and planners must identify relevant uses and activities, which results in behavior-based design (R. Kaplan, 1980; William Hollingsworth Whyte, 1980).

Historically, many studies of landscape architecture have investigated the relationship between humans and the landscape. This was done by developing methods for understanding how people react to the landscape through surveys, observation, and other tools. The strength of big data, especially social media data, is that it is generated by users: users share their thoughts and behaviors with other users in a virtual environment. For landscape architects, understanding people through surveys and social media data can let them link two nodes: human and landscape. However, not all parks have met the needs of the public in recent years, which means that traditional methods may be inadequate or that planners and designers are unaware of other methods of collecting data, such as the use of social media. To examine the value of different ways of collecting data on urban park users' needs, we compare two methods: a survey, representing small data, and social media analytics, representing big data.

Practical problems in designing and planning a park can be explained in two ways: demand and supply. Following the demand approach, the Virginia Department of Conservation and Recreation (DCR) conducts the Virginia Outdoor Recreation Survey of citizens for every five years. This survey has been used to understand people's activities in recreational places and their needs and wants in new facilities. The results become the background of a regional master plan. Another way to understand the needs of visitors is to collect staff feedback. Almost every park has a staff maintaining and managing it. As part of their jobs, park staffs also gather information on the needs and wants of people. A third approach is checking facilities. Every ten years, the DCR adds new facilities on the basis of past use. If a particular facility has been used a lot, more of the same kind are provided. And a final method is using surveys of park visitors to understand their needs and desires; however, these surveys are conducted only for overnight visitors, such as campers.

There are several practical problems with these approaches. First, they may not capture detailed information, such as visitors' activities in and attitudes about a park. For example, the Virginia Outdoor Recreation Survey only asks for general information, such as what people want. It is not designed to collect data for specific parks. Second, they do not provide real-time data. The Virginia survey is conducted only every five years, and the DCR collects staff demands for new facilities only every ten years. Delays in obtaining information may lead to delayed management of a park; real-time data and monitoring are needed. Third, the methods used by the DCR do not collect people's demands directly. The survey questionnaire is designed by the DCR and may not even include real demands. Survey creators must conceive of what things are possible ahead of time. If they don't think of something, it will not be on the survey and thus will not be captured. People who design surveys only try to capture certain things. On the other hand, big data can capture whatever people are motivated to post about. Although there are only some activities people engage in, there are none that won't be posted or captured. On social media, analysts must select the keywords to be used in the analysis. If they miss an important one, they will not capture those data. There is no perfect data set, and this is one reason for complementary studies.

Social media data appear to have potential for addressing the problems described above. This study compares commonly used methods with new kinds of data set. This is motivated by several questions: {1) How do people use linear parks? (2) What are their attitudes and preferences regarding linear parks? and (3) Is there any way to understand park users' behaviors and attitudes? For the first and second questions, designers and planners may use traditional methods such as user surveys. For the third, a new method, social media, is used. To address the practical problems, we investigate the effectiveness and reliability of social media analytics in park studies and its implications for the park planning and management.

3. Research Questions and Objectives

Three major questions arise from the research problem described above, each with its own sub-questions. The first is about gathering information through social media analytics. The second gauges the reliability of social media use by comparing the results of the first question with the results of a survey. The third question examines information that social media fails to capture but surveys can capture.

Research Question 1: What information can landscape architects obtain from a survey?

1-1: Who responses the survey?

1-2: What do they do in linear parks?

1-3: How satisfied are they with their experiences in linear parks?

These questions are designed to identify visitors' attitudes toward and behaviors in linear parks using a survey. They provide detailed information on linear parks and discussions of their use to help us understand park visitors.

Research Question 2: What information can landscape architects get from social media?

2-1: What park visitor activities can be investigated using social media analytics?

2-2: What park visitor interests can be investigated using social media analytics?

2-3: How can park visitors' satisfaction can be investigated using social media analytics?

Research Question 3: How does social media data differ from survey data?

3-1: Who uses social media, and why?

3-2: How are park visitor activities portrayed differently in social media data than in survey data?

3-3: How does park visitors' satisfaction appear differently in social media and in surveys?

The third research question examines the limitations of social media analytics by providing information such as which data social media collects better, which data surveys collect better, and what they can both collect. For example, the demographics, socio-economic status, and education levels of park visitors cannot be collected through social media analytics. By identifying the strengths of surveys and the weaknesses of social media analytics, this study contributes to the identification of appropriate data and analytical methods by park type and their expected results.

Research Question 4: How can landscape architects use social media and surveys?

4-1: In what situation might each type of data be best used?

Overall, the research questions in this study are designed to investigate the effectiveness, credibility, and limitations of using social media analytics in park studies. The first question has implications that can be addressed in social media analytics. The second addresses the reliability of social media analytics. The third describes the limitations of social media analytics by identifying the information missed on that approach.

4. Importance of the Study

This study is designed to help us understand park users through surveys and social media analytics, and it differs in several ways from previous studies. First, this study determines the meanings of social media posts instead of treating them merely as quantities. Social media analytics in park studies have so far taken only a quantitative approach. For example, studies of how many people visit parks and where they check in (Shen & Karimi, 2016; L. Wu, Zhi, Sui, & Liu, 2014; Zhan, Ukkusuri, & Zhu, 2014) have collected and analyzed the numbers of posts people make (Hamstead et al., 2018; Heikinheimo et al., 2017). These studies are meaningful because they use geotagged data to find hotspots during the day and at night, and they reveal new patterns of use in open places. But because they simply count each post as a unit, they overlook the rich information in textual data. By investigating and interpreting these textual data, we expand the use of social media analytics in park studies.

Second, this study addresses rich information about linear parks. While it has been a long time since the High Line opened, and its effects on other settings are now clear, only a few studies have concentrated on who uses these types of park and how they use them. Previous studies have focused on economic benefits such as increased tax income and property value (Ascher & Uffer, 2015; Crompton, 2001; Kang & Cervero, 2009). This study contributes to our understanding of park users by analyzing who visits parks, and why and how they do so, to providing implications for future planning and management of linear parks.

Third, this study contributes to the empirical analysis of the effectiveness of various physical attributes of linear parks. Linear parks on disused railways are characterized by connectivity (via length) and uniqueness (Flink, Olka, Searns, & Conservancy, 2001; Lindsey, Wilson, Yang, & Alexa, 2008). However, there has been no empirical analysis of the effect of these features on the use of the parks. This study analyzes the connectivity and uniqueness statistically with a focus on parks users and their behaviors: for instance, who visit linear parks, why, what they do there, and how satisfied they are.

In summary, this study identifies new ways of collecting information about a new type of park, linear parks, using a new method, social media analytics. The results are then compared with survey results to establish the credibility of social media analytics. Lastly, shortcomings of social media analytics are identified. This study is meaningful in helping us understand the users of new types of parks and suggesting design and planning strategies. Regarding methodology, this study also involves evaluating the use of social media analytics and its advantages, disadvantages, and reliability.

5. Study Organization

This study asks why landscape architects rarely use new-data sets, such as social media data, and how they investigate new types of park, such as those built on brownfields. In the present introductory chapter, we have introduced the reasons for the study, its main idea and research questions, how the objectives will be pursued, and how the study is organized. Although linear parks have been built in many cities, few studies have been done of the behaviors and attitudes of their users, and those mainly through surveys. Big data, especially social media data, may supplement our understanding. The rest of this dissertation is divided into four chapters: a literature review and chapters on methods, results, and conclusions.

The literature review focuses on three subjects: (1) survey data, (2) social media data, and (3) comparisons of the two. We investigate theoretical considerations about the methods used to understand the topic through a review of the relevant literature.

The methods chapter explains the data collection and analytics used in this dissertation. It has five sections. The first describes the research design and its rationale and provides more detail on the three parks selected for this study. The second revisits the research questions and explains how they are addressed. The third section describes phase one of the research, which includes the design of the survey. The data collection, participants, and procedure are explained here. The fourth section describes the design of phase two, focused on social media analytics. The last section explains phase three, the comparison of the two methods.

The results chapter provides the findings on the research questions. It is divided into three sections following the phases of the work. The first section answers research question 1 using a survey. The second answers research question 2 using social media analytics. The last section describes the findings of the comparison study using both methods.

The conclusion chapter summarizes the overall findings and describes their implications for the design and planning of linear parks. These implications include recommendations for planners and designers, governmental agencies, professional societies, and professional education. We also identify an area where more research is needed.

CHAPTER II. LITERATURE REVIEW

1. Social Media Data

“Big data” is a phrase indicating a significant change in data collection, storage, and analysis. Provost and Fawcett (2013) define big data as data sets that are too large for traditional data-processing systems and require new technologies. According to Zikopoulos, deRoos, Parasuraman, Deutsch, Giles, and Corrigan (2013), big data has four dimensions: volume, velocity, variety and veracity. One unique characteristic of such data is every person generates large amounts of in each day; we call this user-generated data. Millions of people use social media services such as Twitter, Instagram, and Snapchat to update others and record their thoughts and locations (Tasse & Hong, 2017). This kind of social media data can help urban planners, policymakers, and landscape architects understand people’s relationships within a city. According to social cognitive theory, users post information and behaviors on social media to communicate their thoughts to each other (Kankanhalli et al., 2005). Big data as a tool for interpreting urban places, and people have the potential to illuminate the relationship between themselves and places in landscape architecture.

In this section, we explain social media and its use in landscape architecture. We first explain what social media is by focusing on its definition, types, strengths, and weaknesses. After that, we investigate social media as a data source and a method of analysis. The social media analytics section describes the structure of social media data, several ways of collecting it, and social media analytical methods. At the end of this subsection, we describe the ways many researchers use social media data and the results they can get.

1.1 Social media platforms

1.1.1 What is social media?

The recent emergence of big data, including social media data, has opened up new possibilities for landscape architecture. Social media involves internet-based applications that provide users with services to create and share content (Boyd & Ellison, 2007; A. M. Kaplan & Haenlein, 2010). Social media has reshaped people's thinking and became an important source of data not only in practical fields but in academic areas such as urban studies, the humanities, geography, and sociology. Social media can improve growth dramatically because of the emergence of "Web 2.0" features (Lin & Geertman, 2019) and the growth in internet access and smartphone usage (Kitchin, 2014). Web 2.0 is a collection of electronic systems that allow users to interact and share information (A. M. Kaplan & Haenlein, 2010). Landscape architecture in particular may be able to use social media data very actively because its purpose is to improve people's everyday environments. Despite that, researchers and other professionals in the field rarely use social media in their work.

Kaplan and Haenlein (2010) classify social media into several categories: collaborative media, such as Wikipedia; media designed to share content, such as YouTube and Flickr; social networking platforms, such as the widely known Twitter and Facebook; and blogging platforms, such as WordPress and Tumblr.

Social media can be used to improve public engagement tools, particularly in two respects: It contributes to greater public engagement, and it acts as a public communication channel. Regarding the former, many researchers agree that social media use enhances public engagement (Kleinhans, Ham, & Evans-Cowley, 2015). It does have population biases, and the skewed age group of the main active users of social media creates limitations. On the other hand, this bias also increases younger generations' participation in urban planning (Schroeter & Houghton, 2011). Fredericks and Foth (2013) coined the phrase "augmenting public participation" to stress the function of social media (Fredericks & Foth, 2013).

Social media becomes a virtual public sphere by providing a supplementary channel to the traditional one. It also provides a new communication space for the public. According to Jendryke et al. (2017), social media can help us understand many issues in a city, such as disasters and emergencies (Jendryke, Balz, & Liao, 2017). Brkovic and Stetovic (2013) also asserted that social media is an important tool for communication within communities. They explain social media platforms as providing outreach capabilities to let urban planners gather information from residents, announce news, and share resources (Brkovic & Stetovic, 2013). Sui and Goodchild (2011) noted the use of social media platforms to improve communication between citizens (Sui & Goodchild, 2011).

1.1.2 Related studies

Researchers use social media data to detect landscapes and land uses, support public engagement, and identify the relationships between people and place. This section introduces work that has used social media as a major data source in urban studies and landscape architecture. Because many studies in landscape architecture use smaller sites than in urban studies, social media use in urban studies is mainly discussed. Many studies have used social media data to detect (1) hidden functions of urban areas and (2) hidden relationships between humans and landscapes.

First, social media helps researchers detect hidden uses of urban areas, green spaces, and protected areas. Tu et al. (2017) discovered urban functions by extracting user behaviors from social media data. They revealed the different uses of urban areas by function, such as work type, and by time (Tu et al., 2017). Similarly, Chen et al. (2017) depicted urban functions using hot-place analysis with social media data. They identified different urban functions and their central locations. Silva et al. (2014) showed that social media data allow the detection of characteristics of each urban area (Silva, Calijuri, Sales, Souza, & Lopes, 2014).

Many studies in urban planning use social media to find functional areas; studies in landscape architecture also use social media to detect ecosystems. Oteros-Rozas et al. (2018) examined more than a thousand photos from Flickr and Panoramio to identify relationships between landscape diversity and ecosystem services. Landscape diversity is an important aspect of the quality of a landscape, and cultural ecosystem services are the benefits green spaces provide to people. The authors found a weak positive relationship between the two factors (Oteros-Rozas, Martín-López, Fagerholm, Bieling, & Plieninger, 2018).

Social media data also provide opportunities to detect soft factors, such as labor market. Batty et al. (2012) asserted the importance of social media data for understanding the labor and housing markets (Batty et al., 2012).

Second, social media provides us with a chance to understand how people interact with the landscape. Tieskens et al. (2018) used photos from Flickr and Panoramio to examine the relationship between landscape attributes and preferences. Many researchers in landscape architecture have focused on the relationship between landscape attributes and people's preferences (Tieskens, Van Zanten, Schulp, & Verburg, 2018), but most studies have used surveys with photographs, and there is a gap between photos as represented media and real places. Studies can reduce the gap between photos and reality by using visitor-uploaded photos.

Multimedia content posted to social media platforms can also be used by researchers to identify new uses of urban place and the relationships between humans and nature. In particular, social media data contain location information that contributes to an understanding of how people use a place. Social media studies that use geo-tagged data often provide valuable information for design and planning. Wu et al. (2017) studied the use of urban trails by examining location information from Flickr and Twitter posts, with the aim of determining the usefulness of social media as a proxy for demand. By mapping geo-tagged posts made along the trails, Wu et al. created heat maps and calculated trail traffic to compare two data sets. They found a weak but meaningful statistical correlation between photos and trail traffic (X. Wu, Wood, Fisher, & Lindsey, 2017). Fisher et al. (2018) also attempted to predict trail traffic using social media. They measured traffic with various techniques, including geotagged photos, and compared the results, finding that the correlation between official statistics and geotagged photos was between 55% and 95%. From these results, they concluded that user-generated data can help urban planners and landscape architects manage and monitor trail use (Fisher et al., 2018).

1.2 Twitter as exemplary social media data

1.2.1 What is Twitter?

Twitter is more widely in the social sciences than other social media platforms. It is a microblogging platform (Shaw, 2017) that lets users post short messages (up to 280 characters) with various options, including check-ins, photos, and videos. As a new data source, Twitter also plays an important role in urban studies. Three features of Twitter in particular make it a popular data source: the volume, quality, and accessibility of its data.

1.2.2 Why is Twitter used in studies?

First, Twitter generates large volumes of high-quality data. Its 330 million monthly active users produce at least 6,000 tweets a second (Twitter, 2019). Twenty to thirty percent of internet users also use Twitter (Blank & Lutz, 2016); in the United States, for instance, one in five adults and one in three youths uses it (Shelton 2016). This makes Twitter a prominent medium for the distribution of information and opinions (Mathioudakis & Codas, 2010).

Second, Twitter provides meaningful content. It enables users to post their own opinions and add various attachments, including photos, videos, and hyperlinks (Shaw, 2017). The data extracted from Twitter also contains textual information intended for many purposes. Whereas Facebook may provide connectivity with peers in a virtual environment, Twitter stands out as a source for the discovery and spread of information and other content (O′Riordan, Feller, & Nagle, 2016). In this regard, it may be useful to trace changes in Twitter over time. In its early years, from 2006 to 2009, Twitter contained mainly useless chatter framed as unworthy of attention. This started to change in 2009, to include informative messages on trending topics (Rogers, 2013). Marwick and Boyd (2011) found one reason for this in Twitter's change of its dialog box prompt from "What are you doing?" to "What's happening?" (Marwick & Boyd, 2011).

Third, public access to Twitter data is more open than with other platforms. Because of privacy problems, many social media platforms block their data from the public. There are ongoing challenges involving social media data from Twitter, Facebook, and similar sites, so that their terms of services are continually being updated and changed to protect privacy (Poorthuis & Zook, 2017). For example, Facebook has increased its restrictions more and more, so at present, people can only access data from specified period and with restricted numbers (Facebook, 2019).

Twitter also restricts data collection, but it provides a longer period (Twitter, 2019) and more content than Facebook. Twitter has several application programming interfaces (APIs), such as streaming API, rest API, and search API (Hawker, 2010), and the streaming API allows the collection of posted tweets in real time (Shaw 2017). This API can also be used without limits, though it does not allow researchers to gather past postings. The rest API enables researchers to access past postings, but it is limited to receiving the number of tweets every 15 minutes (Hawker, 2010). Twitter charges fees to reduce this limitation (Twitter, 2019). One other option for access to Twitter is its search interface, which gives users access to the full history of Twitter (X. Liu, Nourbakhsh, Li, Fang, & Shah, 2015).

2. Comparing Social Media Data and Survey Data

One of the main issues in using social media for design and planning is the validity of social media data. Even though it has become an important resource, researchers may be curious whether its results have explanatory power. To answer this question, some researchers have compared social media analytics with traditional methods such as surveys. In this section, we review the differences between the two approaches. Several studies that have compared the two data sources are also reviewed here. After that, we discuss the use of social media data in landscape architecture. The aim of this section is to provide information on what social media is, how researchers use social media analytics, and how this differs from traditional methods. We thus compare social media analytics, as a new method, with the traditional analytics of the design and planning field. First, we describe the traditional methods, which are frequently mentioned in this study, and introduce their weaknesses. Second, we describe the strengths and weaknesses of social media analytics. After that, we introduce a study of the similarities and differences between the two.

2.1 Surveys as small data

Traditional methods of design and planning fall into two categories: qualitative and observational approaches (Roberts, 2017). Qualitative approaches gather opinions from the user segment. This approach includes any method of understanding users: surveys, interviews, community meetings, and self-reporting (Roberts, 2017). The second kind of approach uses observation at the study site. Field observation has been used in many studies related to urban areas (Birch, 1986; Gehl & Svarre, 2013; Jacobs, 1961).

2.1.1 Strengths of small data

There are several ways to understand how people use a place and interact within it. Researchers measure these variables through field observations, surveys, and interviews. Among these, surveys satisfy two demands: fairness and efficiency. Fairness is related to concepts such as democracy, representativeness, transparency, and public acceptability (Rowe & Frewer, 2005). It concerns participants' and the public's perceptions of whether the research was conducted in a manner that accurately reflects the views of the target population. Efficiency is the ease with which data can be collected (Rowe & Frewer, 2005).

Because many surveys target anticipated park users and gather data about their participants, it easy to verify who the participants are and how representative they are of park users. In many studies, surveys are used to understand people's attitudes and behaviors in urban open spaces such as parks. Peters, Elands, and Buijs (2009) used a survey to find that urban parks can promote social cohesion (Peters, Elands, & Buijs, 2010). Whiting, Larson, Green, and

Kralowec used a survey of park visitors to identify their motivations and preferences in outdoor recreation (Whiting, Larson, Green, & Kralowec, 2017). The present study uses surveys as a representative traditional method to compare with social media data as a means to understanding the activities and preferences of park users.

2.1.2 Weaknesses of small data

By following the growth of the importance of urban areas in the twentieth century, especially urban green areas, researchers in urban studies have developed approaches to assess and manage those areas. Although traditional assessment tools such as surveys, interviews, and observations can be used to evaluate the benefits of parks on-site (Birch, 1986; Carmona, 2010; Ewing, 2001; More, Stevens, & Allen, 1988), a weakness of those methods is that researcher's intentions can be reflected in the questions that are asked and the way they are worded. In contrast to newer social media analytics, those methods are often called traditional methods (Lin & Geertman, 2019; Roberts, 2017). They are well understood and have known strengths for understanding human behavior in certain places (William H. Whyte, 2009) and public opinion about those places. Furthermore, many tools such as SOPARC have been developed for using traditional methods (McKenzie, Cohen, Sehgal, Williamson, & Golinelli, 2006). Although traditional methods do have strengths, they have several limitations in comparison with the new methods: (1) cost and labor intensiveness, (2) staticness in time, and (3) the validity and reliability of the information obtained.

Gathering information using traditional methods requires a lot of money and time. In particular, funding is needed for field investigations such as surveys and observations. Studies of the green spaces in a city require tremendous time and monetary resources when employing traditional methods (Chen, Parkins, & Sherren, 2018). And even when researchers have the fund to conduct a survey, sometimes they fail to find reliable respondents due to a lack of visitors. In these cases, researchers conduct additional online surveys, which require extra finances (D. Wang, Brown, Liu, & Mateo-Babiano, 2015).

Another limitation of traditional methods is their static timeframes. Because these methods only capture data from a certain period, they make it hard to trace changes in the use of a place over time. This may be a reason for the lack of understanding of dynamic use among designers (Chen et al., 2018).

Traditional methods also have problems with the validity and reliability of their information, mainly for two reasons. One is the accuracy of self-reporting. Traditional surveys often ask respondents directly about their opinions, and their answers depend on their subjective interests (Jiang, Li, Larsen, & Sullivan, 2016). The other cause is related to sample size: the size of the surveyed population affects a study's validity and reliability. The samples used with traditional methods are often too small to represent the total population (Cohen, Martires, & Ho, 2016). For example, studies often gather samples of 200 to 1000 people for surveys; by contrast, social media data can have sample sizes above 5,000 or even 100,000.

As for reliability, survey respondents often answer the questions differently depending on the options given to them. For example, respondents will often guess that walking is a more popular activity than running if walking mentioned earlier than running (D. Wang et al., 2015).

2.2 Social media as big data

2.2.1 Strengths of social media

Social media has strengths over traditional methods in the design and planning field in terms of efficiency, volume of data, and variety of data. Regarding efficiency, social media data is known for providing more economical data sets than traditional methods such as a surveys and interviews (Guhathakurta, Zhang, Chen, Burnette, & Sepkowitz, 2019). Due to user privacy concerns, social media platforms have begun blocking free access to social media data (Twitter, 2019; Facebook, 2019), However, researchers can still access social media platforms at little or no cost.

Regarding size, social media platforms expand the accessible boundaries of study area, time, and volume and precision of data, making it possible to track park visitors on finer-grained and broader spatial and temporal scales (Kovacs-Györi et al., 2018). Social media data allow researchers to examine all the parks in a city and count their visitors (H. J. Kim, Chae, & Park, 2018; Kovacs-Györi et al., 2018); indeed, the scale of a study can be expanded from a city to the whole globe (Cheng, Caverlee, & Lee, 2010; Kovacs-Györi et al., 2018; Shelton, Poorthuis, & Zook, 2015). These data also make it possible to trace the dynamic use of a site (Kovacs-Györi et al., 2018). Kim et al. (2019) included all the protected parks among the ASEAN Heritage Parks as their study site. Fisher (2018) selected 16 trails in western Washington. Chen et al. (2018) included all the green spaces in Shenzen, one of the biggest cities in China. As in those studies, social media enables researchers to enlarge their study sites over traditional methods.

A related feature is the possibility of temporal extensions of the study area. Social media data allow extensions of the limited research timespans of traditional methods (Li 2019). Chen et al. (2018) traced the number of visitors to a park over the whole day. Gosal (2019) tracked the use of a site from 2007 to 2016. These kinds of study are impossible with traditional methods.

One other aspect of volume is the fact that social media generates very large data sets. In the case of Twitter, users produce 6,000 tweets a second and 500 million a month (Gosal, Geijzendorffer, Václavík, Poulin, & Ziv, 2019).

2.2.2 Weaknesses of social media

The weaknesses of social media data are related to their strengths, and fall into four categories: (1) population bias, (2) bias toward positive rather than negative information, (3) requested specialty to collect data.

Social media platforms have significant population biases (Ruths & Pfeffer, 2014); though they generate enormous amounts of data, they do not represent the whole population. For instance, Twitter (2019) reported that it has 330 million active users who tweet 500 million times a month. Researchers have estimated that one in five adults and one in three youths in the U.S. uses Twitter. (Smith & Brenner, 2012). At a glance, this makes Twitter look nearly perfect as a medium for representing the population. However, when we scrutinize Twitter users, we find various substantial biases. One is in demographics: Twitter users tend to be male (Mislove, Lehmann, Ahn, Onnela, & Rosenquist, 2011), young (Smith & Brenner, 2012), single, higher-income, and white (Blank & Lutz, 2016), and to live in urban areas (Graham, Hale, & Gaffney, 2014; Smith & Brenner, 2012). Mislove, Lehmann, Ahn, Onnela, and Rosenquist (2011) argued that Twitter users were much more likely to be male and live in high-density areas. They also reported that Twitter users over-represented populous areas and under-represented Hispanics and African Americans (Mislove et al., 2011).

Heavy users create another population bias. According to Kovacs (2018), around 50% of tweets are posted by just 2.2% of users (Kovacs-Györi et al., 2018). Commercial uses of Twitter may also produce population biases: according to Yesmail (2015), 91% of the largest consumer brands had Twitter accounts (Yesmail, 2015). This means that many tweets may contain marketing information rather than users' authentic opinions about themselves. Overall, social media has strengths due to its volume, but it also has population biases.

The second weakness is the accuracy of social media data. A strength of these data is that they are user generated; users can post freely in many forms, using text, hashtags, photos, videos, and URLs. This can be helpful to researchers, as it means they are not taking part in generating the data (Schweitzer, 2014). However, because users can post their thoughts and status freely, social media data tend to contain information the users like, rather than negativity (E. Kim, Lee, Sung, & Choi, 2016). That is, the data are selective and may be connected much more to positive than to negative things.

Another accuracy problem is related to location information. Because geo-tagging in social media data lets researchers uncover check-in information over large areas, it is used in many studies (Evans & Saker, 2017; Heikinheimo et al., 2017; Kelley, 2013). But because this information can also be generated by users, it not match users' actual locations (Stefanidis, Crooks, & Radzikowski, 2013). Furthermore, users may remove their locations for their own benefit (Elwood, Goodchild, & Sui, 2012). Thus social media has strengths for including user-generated data, but this also involves limitations such as selective data and fake check-ins.

The third weakness to using social media involves accessibility, in two areas: data collection and data interpretation. To collect data, researchers must know a programming language to use platforms' APIs—interfaces that lets third-party developers create data sets (Lomborg & Bechmann, 2014). To analyze their data, researchers must also know programming languages such as Python, C++, and Java. Mannila et al. (2006) compared the difficulties of programming languages and reported that Python is easier to learn and use than Java (Mannila, Peltomäki, & Salakoski, 2006). However, learning any new language is difficult for a beginner.

Another accessibility problem is the difficulty of extracting valuable information from social media data. First, researchers can often use just one programming languages to extract useful information from social media data. Although several agents such as Sprout Social, Awario, and TapInfluence offer services for using social media data more easily ("11 of the Best Social Media Analytics Tools for Brands," 2019), they provide only simple analytics that are not suitable for studies. And even when researchers can use programming languages and know how to collect data, they must handle the data so as to reduce biases and extract information with related techniques.

2.3 Comparative study

Li (2019) compared surveys and social media data as a future research methodologies for landscape architecture. Li explained the strength of traditional surveys as follows: (1) Surveys don't rely on the accuracy of the device. (2) The extracted information tends to represent the population, (3) Data collection and analysis tend to be easier than with big data. (4) Researchers can directly contact users to understand them better (Li, 2019). Heikinheimo (2017) also stresses the accuracy and quality of the information as advantages of the traditional approach (Heikinheimo et al., 2017).

Several urban studies researchers have compared traditional and social media analytics empirically. Wang, Jin, Liu, Li, and Zhang (2018) compared social media data and survey data to assess the attractiveness of a park. They extracted data from two major social media platforms in China, Dazhongdianping and Mafengwo, and collected 5,440 posts from April 2013 to October 2015. Then they conducted an online survey asking park visitors about the attractive items in the park and collected 243 responses from 10 April to 15 April, 2016. To compare the attractiveness judgments from the two sources, they divided the social media posts into positive, neutral, and negative and identified keywords from each sentiment group. After that, they analyzed the survey and classified attractiveness factors into core attractions, important attractions, and marginal attractions, then compared the ordering of the attractiveness factors (R. Wang, Zhao, & Liu, 2016). The findings from the social media data and survey data overlapped, they found, and reflected the attractiveness of the park well. The authors also pointed out several differences: (1) A large volume of social media data does not predict the quality of the data. (2) Social media provides unstructured data. (3) Surveys are well suited to gathering valuable information. They concluded that social media and survey data complement each other by correcting each other's weaknesses (R. Wang et al., 2016).

This study provides meaningful implications for future studies by comparing traditional methods with social media analytics empirically. The authors tried to control for population biases by selecting age groups for the survey based on actual social media users in China. Because it is difficult to determine social media users' demographics, this approach may have reduce bias. However, age is a factor in population bias—many studies list age, gender, income, education level, living area, and density of the area—and selecting specific age groups may have created another bias. The study explicated the attractiveness of parks with a focus on young people rather than the general public, and the authors dismissed the temporal differences between the two methods. The survey was conducted in April 2015, and the social media data included tweets from 2013 to 2015. Because park visitors and their perceptions of the park may be related to the season, the authors had to control for seasonal differences. Although this study may have limitations, it was a worthwhile attempt to compare social media data with surveys through an empirical case study.

3. Summary

Table 1 shows the strengths and weaknesses of two data sets, one from a survey and one from social media analytics, based on the previous study. According to some studies, surveys advantages in fairness and efficiency. With a large population, surveys are more efficient than individual or focus group interviews. However, surveys also have low accuracy in self-reported information and can require considerable costs and labor to conduct.

On the other hand, social media provides large volumes of data in a variety of formats, such as text, images, and videos. Social media data can also be collected and analyzed more easily and effectively than survey data. However, social media is used more by younger people than by the elderly. Because of that, it has a population bias. Social media users also tend to post more positive than negative information. Finally, a major limitations to collecting social media data is the need to learn a computer language.

Table 1. Comparison of social media analytics and surveys.

	Survey	**Social Media Data**
Strengths	• Fairness (Rowe & Frewer, 2005) • Efficiency (Rowe & Frewer, 2005)	• Efficiency (Guhathakurta et al., 2019). • Volume of data (Kovacs-Györi et al., 2018). • Variety of data (Li, 2019).
Weaknesses	• Lack of accuracy in self-reported information (Jiang et al., 2016). • Limited sample size (Cohen et al., 2016). • Costly and laborious.	• Population bias (Ruths & Pfeffer, 2014). • Positive rather than negative information (Schweitzer, 2014). • Programming language needed for data collection (Lomborg & Bechmann, 2014).

Table 2 shows the characteristics of data sets used in planning and designing a park. It compares a survey and social media analytics in terms of descriptive data, time, and users, based on previous studies. Surveys have been considered a valid and reliable data source (Groves, 2004). Although a survey can capture only the present, it is a useful tool for collecting pre-set answers. On the other hand, social media analytics can cover past and real-time data (Ahmad, 2018; Goodspeed, 2013). In terms of validity, social media data are weaker than surveys, which have been used for a long time.

Table 2. Characteristics of data sets used in planning and designing a park.

	Criteria	Survey	Social Media
Descriptive data	Validity	●	○
	Reliability	●	●
Time	Past	N/A	●
	Present	●	●
	Real-time	N/A	●
User	Needs or desires	○	○
	Satisfaction	○	○

CHAPTER III. METHODOLOGY

The purposes of this study fall into two main areas: social media analytics as a methodology in park studies, and linear parks as a new types of park, especially those built on unused railways. Regarding the first, we propose (1) to identify the information that social media can gather from textual data, (2) to examine the reliability of social media, and (3) to identify information that cannot be collected through social media, in order to understand its limitations. Regarding linear parks, the aim is to provide and implications for future planning and management by answering several questions: (1) Who uses linear parks, and how? (2) How do the characteristics of linear parks affect their use? 3) What data are required for which linear parks?

This chapter has four sections. The first is about the pilot research and explains why researchers collect textual data from social media sites or choose to use surveys. The second section describes the overall research design, introducing the analysis to be done and the three targets, and methodologically explaining the research questions. Sections three and four describe the data collection and analysis methods, categorized by data type; the third section describes how social media data were collected and analyzed, and the fourth section describes how the survey data were collected.

1. Pilot Test for Data Selection

The pilot test was conducted at Gyeonggui Line Forest Park in Seoul, Korea, which opened in 2016. It is built on disused railways and emerges in an urban hot place (Seoul Metropolitan Government, 2018). This park was selected for several reasons: (1) its background, built on an old rail track; (2) its popularity: the park became an urban hot place within a short time; and (3) familiarity: it was easy to collect and understand Korean data. This park was suitable for collecting both big and small data. For big data, tweets posted from June to September 2018 were collected using the Twitter API and including geo-locations. A total of 3,703 tweets and 177 responses were collected. For small data, various methods were used: Field observations and surveys were conducted in the park, from August to October 2018; and interviews with the designer of the park and nearby residents were conducted on August 20.

There were several findings from the pilot test. First, a survey provides a data set comparable with those of big data. The interviews with the designers and residents provided valuable information on the main design theme and the actual uses of the park. However, for a comparative study of big and small data, this information was not enough. For example, one designer explained the park design process and the conflicts that arose with residents before construction. But because the designers are not involved in the management of the park, they do not have information on its actual usage. The interviews with residents were also good for collecting their opinions on the park, but the sample was too small to compare the results with either the survey results or the social media data. For these reasons, a survey as small data and social media data as big data were selected for the main study.

Twitter's API is a good way to collect all the metadata of each posting, but it has limitations. For example, it offers no access to past data; for past data, one must pay according to the volume of data, response times, and date (Twitter 2019). Even though we collected data through the Twitter API, not all the metadata were needed in the pilot test. For example, location data were collected but not used. This is because the number of tweets that include location information was too small: fewer than 30 tweets, or less than 1% of the total. When researchers want to track people's geographic distribution, location data may be useful, but when researchers are collecting social media data related to a certain topic or place (through keywords searching), it is better to exclude these data.

From the pilot test, two main findings contributed to this study. The first was the need to exclude interview when comparing results with big data. The second was the need to exclude location metadata when collecting large data sets.

2. Research Design

2.1 Research framework

This study is designed to examine two methods, surveys and social media analytics, for three purposes. The first is to understand who uses linear parks and how they do so, via a survey, to develop implications for the planning and management of linear parks in contemporary landscape architecture. The second purpose is to learn about the efficiency and applicability of social media analytics in planning and managing parks. For this purpose, textual data were collected from Twitter and analyzed using text-mining methods. The third purpose is to compare the two methods to determine the usefulness of social media analytics in park studies and to finding an appropriate method of studying linear parks. The study is divided into three phases focused on different methodologies: (1) surveys and statistical analyses, (2) social media data and text mining, and (3) comparison of these two. Through these three phases, the study provides practical guidelines for selecting data and analytics for each type of linear park.

In phase one, we investigate park visitor behaviors and attitudes through a survey. The responses are analyzed by using statistical analyses such as PCA, correlation, and ANOVA to address both linear parks' visitor characteristics and the effectiveness of surveys. In phase two, social media analytics are used to investigate the same facts. The object is to use social media data for planning and managing linear parks(Figure 1). Although these differ from other parks such as neighborhood parks, amusement parks, and national parks, few studies have focused on linear parks' physical attributes instead of on their users.

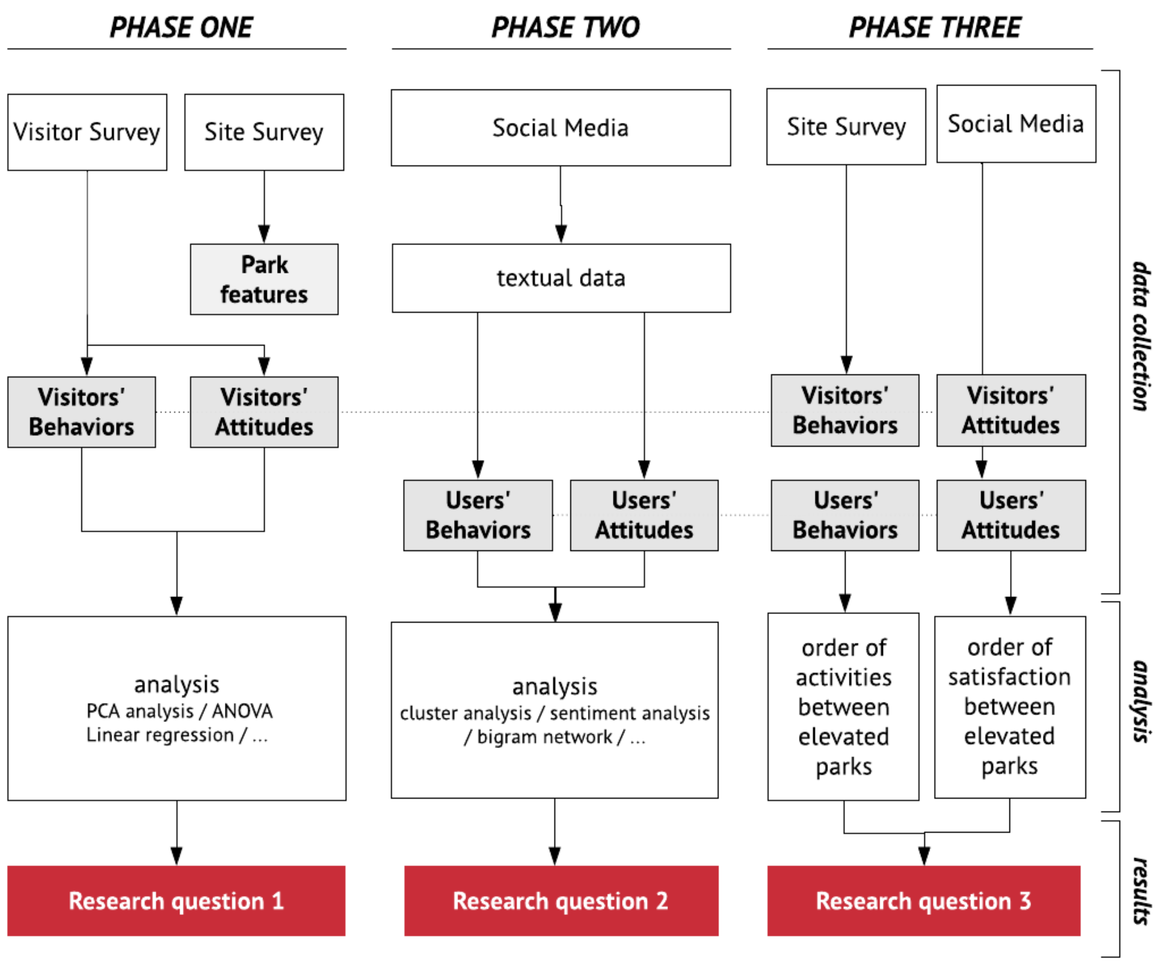


Figure 1. Research Framework

Table 3 summarizes the variables and analytics. This study uses four variables: (1) park visitor behaviors, (2) park visitor attitudes, (3) social media user behaviors, and (4) social media user attitudes. Park visitor behaviors were collected using the survey and divided into two sub-variables: motivations and activities. Park visitor attitudes toward linear parks were collected in terms of the benefits they perceived in their visits and their levels of satisfaction with park features. These features were identified from the literature, mass media sources such as news, and site surveys. Three substantial features of linear parks are connectivity, continuity, and uniqueness. Park visitors' behaviors and attitudes and park features are analyzed using statistical analytics, such as PCA, ANOVA, and linear regression.

Textual data posted to Twitter were also collected to analyze users' behaviors in terms of activities, and their attitudes in terms of sentiments and opinions. These textual data were analyzed using text mining techniques such as sentiment analysis, topic analysis, and bigram networks.

Table 3. Variables and Analytics

Data Source	Variables	Measurements		Independent or Dependent	Analytics
Survey	Visitor satisfaction	Benefits	• Natural • Social • Health	Dependent variable	• PCA • ANOVA Regression • Etc.
	Visitor activities	Activities	• Physical • Social • Relaxation • Viewing	Independent variable Dependent variable	
Social media	User interests	Sentiments	• Positive • Negative • Neutral	N/A	• Text mining • Sentiment • Topic analysis • Clustering • Etc.
		Opinions	Topic distribution	N/A	
		Satisfaction	By activity	N/A	
	User activities	Activities	• Physical • Social • Relaxation • Viewing	N/A	

2.2 Study sites

Three linear parks were chosen for this study, for the following reasons. First, they have similar backgrounds, opening dates, and purposes: all three were built on unused railroads and opened between 2009 and 2015 to provide green spaces to the public (Table 4). Second, all three have unique characteristics in terms of location, attributes, users, and functions. For example, one provides a recreation place, another offers a space for the arts, and the last allows visitors to be in contact with nature. By reviewing and comparing the three, we can determine what kinds of data are valuable for each type. Third, many studies have examined the economic benefits of replacing disused railroads with linear parks, but they have not investigated other aspects of these parks, such as who visits them and what they do. Many researchers have investigated the High Line and its economic impact on its city (Ascher & Uffer, 2015; Levere, 2014). Regarding the 606, some researchers have examined its design, but not its users or their behaviors. And the High Bridge has been mentioned only by media outlets to announce its opening; it has never been the subject of a study. Fourth, each park is located in a different context, which allows us to compare them and find appropriate data sets for their different situations. The 606 is located in a residential area, the High Line is at the center of a commercial area, and the High Bridge is in a rural area.

Table 4. Study Sites

	The 606	The High Line	The High Bridge
Location	Chicago, IL	New York City, NY	Farmville, VA
Ownership	Chicago Park District	City of New York	Department of Conservation and Recreation, Virginia
Average annual visitors	1,000,000	7,000,000	200,000
Opened year	2015	2009, section 1 2011, section 2 2014, section 3 2019, final section	2008, east part 2009, west part 2012, High Bridge

For these four reasons—similar situations, different purposes, investigations into their benefits, and distinct environmental context—these three parks were selected to help us understand (1) what visitors do in linear parks, (2) what they evaluate in linear parks, and (3) how satisfied they are with their experiences. We use social media analytics to (1) provide rich information on linear parks, (2) investigate the pros and cons of social media analytics, and (3) promote the use of social media analytics.

The 606

The 606 was formerly the Bloomingdale line, a linear railroad crossing Chicago from east to west (Figure 2). The railroad was constructed in 1873 and elevated in the 1910s to reduce pedestrian fatalities. It was used for passengers trains but ceased operation in 2001. From 2002 to 2004, the greenway idea was introduced to the public, and the Friends of the Bloomingdale Trail was launched in 2003 to advocate the use of the railway as a park (Gobster, Sachdeva, & Lindsey, 2017). In 2015, the 606 opened as the longest elevated park in the U.S.

Figure 2. The 606: (a) facilities, (b) observation point, (c) pathway.

The design objectives and features of the 606 are as follows: (1) enhancing the unique attributes of the Bloomingdale rail line, (2) balance trail and park, (3) consider residential areas' privacy, (4) provide integrated accessibility for all kinds of transportation in the city, and (5) allow spaces for the arts (City of Chicago, 2004). For the first of these, the design of the 606 aims to preserve and capitalize on the feel of the Bloomingdale rail line by using the existing massive concrete structure but filling it with soil and restoring the industrial infrastructure. The design team focused on these features and maintained the structure to reflect the history of the line (City of Chicago, 2004). To balance trail and park, the 606 has separate bike and walking paths, shared-use paths, and observation spots with seating. For safety concerns between pedestrians and cyclists, the designers were careful to add physical barriers. Because the 606 cuts across residential areas, separation of public and private areas was also an important issue. The designers preserved the privacy of adjacent areas through plants and the linear structure. For accessibility, thirteen access points to the 606 were built. The designers also considered the network of existing transportation and selected the access points carefully.

The High Line

The city of New York allowed the construction of the street-level railroad tracks along Tenth and Eleventh Avenues on Manhattan's West Side in 1847 to ship commodities such as dairy products and beef into lower Manhattan. Numerous safety problems arose, and the area came to be known as "Death Avenue" (Friends of High Line, 2019). To improve matters, in 1929 the city, the state, and New York Central Railroad initiated the West Side Improvement Project, led by Robert Moses (Bighorse, 2010), and the tracks were turned into an elevated railway. Because the line ran through the middle of several blocks, 640 buildings were demolished (Friends of High Line, 2019), and the construction split many communities in two.

The design strategy of the High Line has three main features: (1) The proportion of paving to vegetation, (2) the vegetation, and (3) a memorable design for the railway. For the first, field operations provided several prototype paving arrangements along the line, from 100%

Figure 3. The High Line: (a) facilities, (b) observation point, (c) pathway.

paved to fully unpaved (100% vegetation). Because the High Line is long and narrow, if the same pattern is simply repeated, users may grow bored. This strategy provides opportunities for varied experiences. It also overcomes some limitations of linear parks, such as shallow soil and a dry surface (Fehrenbacher, 2014).

The design of the vegetation itself aimed to maintain the area's current biodiversity and sustainability by using a self-seeded landscape (Figure 3). In the 52 years after the line ceased operations, it was overgrown by native vegetation. Piet Oudolf, the garden designer who lead the planting design of the High Line, says it was inspired by many natural processes and is intended to create a variety of moods that evoke patterns of nature (Friends of High Line, 2019).

The third feature is a memorialization of the history of the tracks. A couple of strategies are employed for this. A peel-up design language is used for benches, bike racks, planters, and a fountain. The paving pattern also represents the rail tracks, and some tracks were retained to let visitors know where they are.

One last feature is the overlooks. There are many lookout points on the High Line, from its endpoint in the meatpacking area to the theatre, offering numerous vistas for visitors to enjoy the city. Overall, the design features of the High Line provide diverse experiences.

The High Bridge

The High Bridge is a part of the High Bridge Trail State Park in Virginia (Figure 4). This park is a 31-mile trail on land that was once part of the Norfolk Southern Railway, which operates 19,420 miles of route in total (Norfolk Southern, 2018). After Norfolk Southern's last train stopped running on this line in 2004, the company donated 31 miles of railroads to the Virginia Department of Conservation and Recreation (DCR), who included the trail in the Virginia State Park system. In 2012, the DCR opened the park to provide recreational services to the public (Virginia Department of Conservation and Recreation, 2019).

Figure 4. The High Bridge: (a) facilities, (b) observation point, (c) pathway

The High Bridge Trail State Park opened in 2012, but the DCR established its objectives for the park in 2006. According to its report, the purpose of the park was (1) to afford non-motorized pathways for pedestrians, cyclists, and horseback riders, (2) to protect historical and natural resources along the trail, and (3) to make the region's history available to individuals, communities, and the country. In 2012, the DCR identified four experiences of visitors: rural remote, rural social, urban social, and a focal point. The focal point, the High Bridge, is the highest part of the trail (DCR Summary, 2012).

3. Research Questions

This section details the research questions and sub-questions listed in the introduction.

3.1 Research question 1: What information can landscape architects obtain from a survey?

This question refers to park visitor surveys, which have long been widely used by landscape architects. It has three sub-questions:

1-1: Who responds to the survey?

1-2: What do they do in linear parks?

1-3: How satisfied are they with their experiences in these parks?

The second of these is about users' activities in linear parks. These parks have unusual features, being longer and narrower than other parks. By answering this sub-question, planners and designers can understand why people visit linear parks and how they use them. Park users' behaviors have two aspects: motivations and activities.

The third sub-question is designed to gauge users' satisfaction with linear parks. Satisfaction is considered a substantial factor in behavior studies. Empirical studies of satisfaction have identified two main issues: (1) providing feedback to staff and (2) providing indicators of experiences and their outcomes (Floyd, 1997).

3.2 Research question 2: What information can landscape architects obtain from social media?

The objective of this question is twofold: (1) examine the possibility of using social media analytics, especially with textual data, to understand users' behaviors and perceptions, and (2) track past data to study the transformation of park usage by social media users. This question has three sub-questions:

2-1: What park visitor activities can be investigated using social media analytics?

2-2: What park visitor interests can be investigated using social media analytics?

2-3: How can park visitors' satisfaction can be investigated using social media analytics?

The first of these will capture visitors' activities in linear parks by extracting keywords representing activities. The second will help us understand how social media users post about linear parks, and whether they view those parks positively or negatively. From January 2015 to July 2019, data were used to track how social media users reacted to linear parks on Twitter. The last sub-question explores the relationship between social media users' park activities and their satisfaction, treated as either positive and negative. If the objective of research question 1 is to identify unique characteristics and uses of linear parks, the objectives of research question 2 is to find an appropriate way to use social media data to understand its users' behaviors and perceptions.

3.3 Research question 3*:* How does social media data differ from survey data?

This question compares surveys and social media analytics to examine the value of the latter for designing and planning linear parks. Although social media data have the potential to illuminate park visitors' activities and perspectives, they have rarely been used in landscape architecture until recently. By comparing social media analytics to an accepted and reliable technique, we can investigate its reliability and uniqueness. This question has two sub-questions:

3-1: Who uses social media, and why?

3-2: How are park visitor activities portrayed differently in social media data than in survey data?

3-3: How does social media data on park visitors' satisfaction differ from survey data?

The first of these is about social media users: who they are and why they post on these platforms. These questions were asked in a park visitor survey. The second sub-question is about the activities of linear park users and may help planners and designers understand why people visit these parks and how they use them. Park users' behavior has two elements, motivations and activities. The third sub-question is designed to assess users' satisfaction with these linear parks. Satisfaction is considered a substantial factor in behavior studies. Empirical studies of satisfaction identify two main components: (1) providing feedback to staff and (2) providing indicators of experiences and their outcomes (Floyd, 1997).

One of these sub-questions is about people's activities, and another is about their attitudes. Because the park visitors who completed the survey are not the same people who posted on social media about linear parks, it's difficult to compare the survey results and the social media results. Instead of doing so directly, I conduct an order comparison. For instance, I compare the parks' order of popularity for physical activity, as given by the survey, with their order in terms of social-media mentions of several keywords related to physical activities. Thus if 30% of visitors to park X and 20% of visitors to park Y did physical activities, the order of physical activity is X, then Y. Comparing the two methods in this way helps to understand the pros and cons of using social media analytics in landscape architecture.

3.3 Research question 4: How do landscape architects use social media and surveys?

This question addresses the practical implications of the who and why questions. Where questions 1 and 2 studied social media and survey data, this question studies social media users and survey respondents to gather practical findings. It has one sub-question:

4-1: In what situation might each type of data be best used?

This sub-question provides valuable information by sorting types of linear parks for practitioners in landscape architecture.

4. Social Media Data

4.1 Data collection

Twitter was chosen for social media analytics for several reasons: (1) Twitter users upload texts and many other types of media, such as photos. (2) Twitter is convenient to access and collect data from. (3) Many studies have used Twitter as their main data source. Following the approached used by Liu (B. Liu, 2015), I codified keyword-based queries related to each park and then submitted each keyword to Twitter's search interface to collect full tweet histories, instead of using the API. I then used a web scraper to collect the tweets automatically. Tweets posted from January 2015 to June 2019 were collected for each linear park (Table 5).

Table 5. Total number of tweets

	Keywords	Number of tweets before pre-processing	Number of tweets after text cleaning
606	"The 606," "Bloomingdale trail"	14,340	12,952
High Line	"The High Line"	206,229	165,347
High Bridge	"High Bridge trail"	457	457

Because most text-mining analytics, including sentiment analysis, are developed for English, using non-English data in the analysis produces inaccurate results (Z. Wang, Joo, Tong, & Chan, 2014). To reduce this inaccuracy, I include only English tweets in the data and exclude others. From tweets related to the 606 after the initial keyword search (n = 14,340), I delete 497, leaving only English tweets (n = 13,843). From the High Line tweets (n = 206,229), I delete

81,000 (39.28%). For the High Bridge, all tweets were posted in English.

After being pre-processed for text mining, the data were reviewed by park and unrelated tweets were removed to reduce inaccuracy. First, the number of tweets per day were plotted for each park to reveal abnormal dates. All tweets posted on those dates were reviewed, and keywords were used to delete unrelated tweets.

Second, heavy users—those who posted more than average—were identified, and their user introductions on Twitter were reviewed for unrelated tweets. For example, a park's name might be that of a business in another: there is a jazz bar called the 606 Club in London, and there is a Scania truck model called the Highline. Through this filtering process, tweets unrelated to the linear parks were deleted.

In the case of the High Line, I identified the top ten dates for tweets and found a special event on each. For example, on January 7, 2016, Singapore announced a plan for a linear park and mentioned the High Line to explain it. Seoul on May 14, 2015; Sydney on August 27; and San Francisco on June 21 also mentioned the High Line when announcing linear park projects. Although these tweets include the keyword, "the High Line," they were not related to the use of the park and were deleted. Next, I identified the top ten heavy users and reviewed their tweets. One heavy user used "High Line" as a product name; this user's tweets (n = 6026) were deleted. The total remaining tweets for the High Line number 165,347.

The 606 data are much simpler than the High Line data. Ten abnormal dates were identified and reviewed; they were all related to events or crime on the 606 except one (October 25, 2016), which was the fortieth anniversary of the 606 Club in London. Heavy users were also reviewed; one account that posted 130 tweets belonged to the 606 Club in London. Tweets including the keywords "the 606 Club" were deleted. The remaining tweets number 12,952.

The High Bridge Trail Park is in a more rural area than the others and has fewer visitors. From 2015 to 2019, only 457 tweets were posted about it, all in English. Abnormal dates and heavy users were reviewed, but all the tweets were about the High Bridge Trail, so the total remains 457.

The pre-processed textual data were then cleaned as follows: (1) stopword removal, (2) lower casing, and (3) normalization. The Natural Language Toolkit (NLTK) in Python was used to remove stopwords. Bird, Klein, and Loper (2009) developed this program for statistical natural language processing for English (Bird, Klein, & Loper, 2009). I also converted all uppercase letters to lowercase and normalized tweets to prevent errors.

4.2 Data analysis

I conduct three main analyses: (1) sentiment analysis, (2) topic analysis, and (3) network analysis. Sentiment analysis is intended to detect public attitudes toward the linear parks (Figure 5). Using the VADER model (Hutto & Gilbert, 2014), I assign each post a sentiment score and sentiment type: negative, neutral, or positive. For clustering analysis and network analysis, I tokenize each post using the bag of words model. This model divides a textual corpus into units and gives a vector number to each. This allows researchers to detect term frequency and relationships between corpuses. The aim of the topic analysis is to identify the main topic each month to tracing how people use linear parks over time. Topic analysis lets us understand the main themes of textual data. Network analysis helps us identify and represent relationships between the main idea and other issues and events discussed on social media. Through network analysis, I can visualize these relationships.

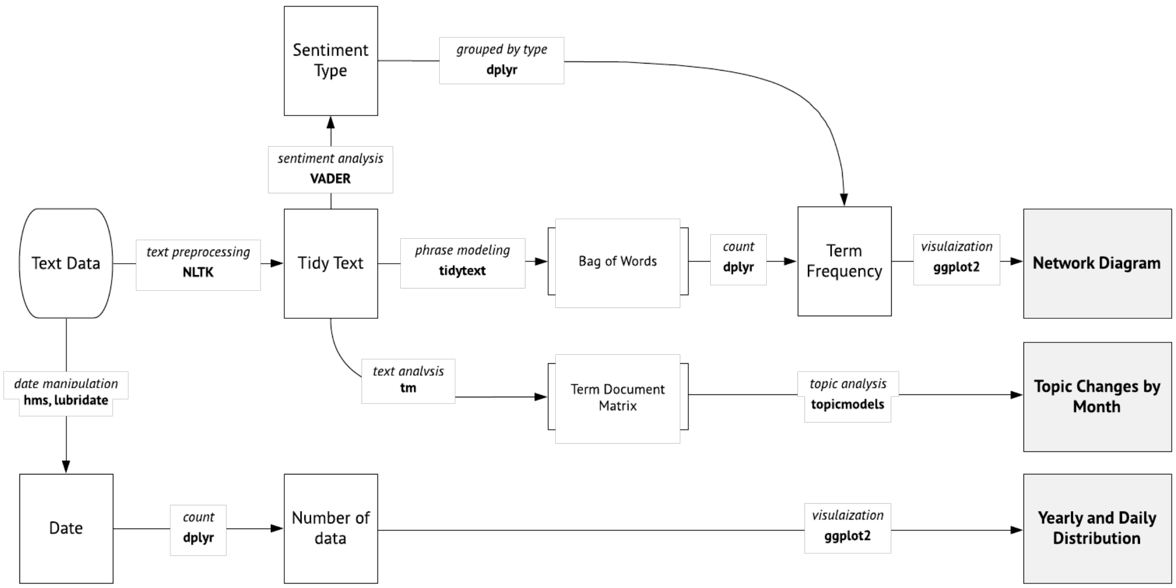


Figure 5. Flow chart of social media analytics.

This study uses social media data to detect public sentiment about and use of linear parks. For the first purpose, I conduct a clustering analysis to detect the terms most frequently mentioned when users post thoughts about linear parks. After measuring term frequency, I generate a word cloud to visualize the differences between three parks. Using clustering analysis, I then group terms to investigate their hierarchical relationships.

For sentiment analysis, I use the VADER model, which was developed especially for social media data. This model provides polarity sentiment types and an overall score based on each corpus's score. If a score is less than 0, the sentiment type is negative; if greater than 0, the sentiment type is positive. A score of 0 means neutral sentiment. VADER was chosen for its accuracy, especially with short texts, relative to other sentiment analysis models.

Next, I parsed each post into a bigram. This is a bag of words tool that divides textual data into two sequential words, labeled "word 1" and "word 2." After that, I conduct a network analysis to check the term frequency of each sentiment. The aim of these analytics is to identify differences between negative and positive sentiments.

Another way to analyze social media data is searching keywords to identify users' activities in linear parks. The previous analytics identified some keywords and relationships between them; this analysis filters the data using the keywords to compare sentiments. Keywords related to physical, social, nature-friendly, overlooking activities are selected on the basis of the survey questionnaire, and then the sentiments about each activity are compared. This can reveal the relationships between activities and sentiments in the virtual environment.

5. Survey Data

5.1 Survey questionnaire

5.1.1 Pilot test for survey

A pilot study using a survey questionnaire was conducted at a landscape architecture studio at Virginia Tech in April 2019. Undergraduates, graduate students, and faculty members in landscape architecture or architecture completed the questionnaire and then checked its feasibility. The aim of the pilot study was (1) to develop a standardized survey questionnaire for use at multiple linear parks, (2) to evaluate the validity of the questionnaire, and (3) to examine the questionnaire's length and completion time. Overall, 63 surveys were collected. Based on the collected data and the opinions of participants, the questionnaire was revised to clarify the main concept of each question and to improve data collection protocols. For example, several answer options were added to question 10 to clarify it. Sections on daily use of social media platforms were also added to gather information on park visitors' social media activities.

5.1.2 Survey questionnaire

The survey questionnaire is designed to address the following factors: trip attributes, activities, motivations, satisfaction, perceived benefits, and user characteristics. The first section includes variables related to the trip attributes. Travel patterns are important for figuring out why users went to a park, how they arrived and how long it took, and who they went with. Six questions ask are used to identify trip attributes for each visitors. The importance of trip attributes is well established in recreation studies (Betz, English, & Cordell, 1999; Keith, Larson, Shafer, Hallo, & Fernandez, 2018), as they play a major role in people's decisions about recreational activities. In park design and planning, trip attributes are also important for anticipating users' activities.

The second section asks about the activities at each park. Most of the activities included were identified from literature reviews (Ahern, 1991; Keith et al., 2018; Whiting et al., 2017); among them are physical, stress-relief, natural, and social activities. The only newly added option is social media activity. Although social media use is widespread in urban spaces (Ahmad, 2018; Chou, Hunt, Beckjord, Moser, & Hesse, 2009), many studies overlook its use in parks. The survey asks users why they post their activities to social media.

The third section asks about the motivations for visits. Outdoor recreational activities are influenced by the variables that move people to take part in them (Whiting et al., 2017), including social contact, relaxation, experience of nature, and physical health (S.-H. Lee, Graefe, & Li, 2007; Whiting et al., 2017; Williams, Schreyer, & Knopf, 1990). These variables are evaluated using a 5-point Likert scales as follows: 1 = not at all important, 2 = slightly important, 3 = moderately important, 4 = very important, 5 = extremely important.

Understanding people's motivations and actual activities at each type of linear park can help us segment the attitudes and behaviors of the participants in various activities. Investigation of the relationship between attitudes and behaviors has led to the recreation opportunity spectrum (ROS) and the recreation experience preference (REP) scales (Beverly L. Driver, 1977; Beverly L. Driver, Brown, & others, 1978). These methods provide urban planners and landscape architects with an array of recreation opportunities to consider when designing a park.

The fourth item on the questionnaire involves preferences about the park. Following the characteristics of linear parks identified in chapter 2, the survey asks about preferences and satisfaction regarding accessibility, connectivity, overlooks, nature, historical features, facilities, and amenities. A 5-point Likert scale is used as follows: 1 = very unsatisfied, 2 = unsatisfied, 3 = neutral, 4 = satisfied, 5 = very satisfied. The Likert scales in this study have the same format (e.g., from disagree to agree) to minimize confusion in respondents. Preferences regarding parks are important to identify the factors most valued by users.

After that, the perceived benefits of the park are asked about. Whereas motivations lead users to pursue activities in a park, the activities help users realize desired benefits. In this study, perceived benefits are classified into three types: social, stress-reduction (B. L. Driver, Brown, & Peterson, 1991; Ulrich et al., 1991), and physical fitness (Godbey & Mowen, 2011; Stodolska, Shinew, & Li, 2010). Understanding perceived benefits is necessary to making decisions about park management (Whiting et al., 2017). These questions also use a 5-point Likert scales, as follows: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.

The final section collects demographic variables such as the user's gender, age group, race or ethnicity, education, and income. Many studies have investigated differences in physical activity preferences according to these traits (Floyd, Spengler, Maddock, Gobster, & Suau, 2008; Whiting et al., 2017).

5.2 Data collection

5.2.1 Survey instrument

This section explains the survey protocol. An intercept survey was conducted on-site. Intercept surveys have been accepted as a reliable and valid data collection method in other trail-based studies (Troped, Whitcomb, Hutto, Reed, & Hooker, 2009). The survey was conducted from April 18 to May 12, 2019, in the linear parks. Sampling dates were selected to include at least two weekdays and two weekend days. To collect more data, an extra survey was conducted for the High Bridge Trail in Farmville from May 18 to 19, 2019.

In each linear park, I approached every fifth adult visitor and asked them to take a voluntary survey on the park. I also outlined the study's purpose and procedure, as follows:

> This study investigates the relationship between park visitor behaviors and attitudes. This survey will take 5 to 10 minutes. This survey has no more than the minimal risk in participating, and you may stop participating at any point without penalty or loss of benefits. By participating in this survey, you may have a chance to understand what landscape architects do for managing a park and contribute to providing information to the public to guide future policy.

When visitors wanted to participate, I provided a copy of the questionnaire on a clipboard with a pen and waited to answer any questions they might have or assist them with any difficulties. For example, due to the weather in Chicago, a couple of participants there asked about the average number of visits and whether they should consider the whole year or only the spring and summer months. After visitors completed the survey, I thanked them for their time and effort and their contribution to the research. In all the parks except the High Bridge Trail, I collected an average of 5 to 6 surveys per hour.

Table 6. Response rate and Sample size

	The 606	The High Line	The High Bridge	Total
Fully response[1]	63.12% (n = 320)	42.18 (310)	84.51 (120)	54.19 (750)
Partial response[2]	1.58 (8)	1.63 (12)	1.41 (2)	1.59 (22)
Refusal	35.31 (179)	56.19 (413)	14.08 (20)	44.22 (612)

[1] Total sample size for the study.
[2] Disqualifying responses.

5.2.2 Response rate

The overall survey response rate was 54.19%, but it differed between the parks (Table 6). The lowest rate was 42.18% on the High Line, due mainly to language barriers: many visitors came from other countries and didn't speak English, and most of these refused the survey. Most visitors to the High Bridge Trail participated gladly, but that park's total visitors (<80 on a weekday) were far fewer than the other parks (>400). A total of 750 responses were collected, excluding partial responses. Respondents who skipped more than five questions were excluded.

5.2.3 Approvals and human subjects review

Before the survey was conducted, the questionnaire and survey instrument were approved by the Institutional Review Board (IRB) Office of Virginia Tech. The approval letter is in Appendix A, the survey protocol in Appendix B, and the approved survey form in Appendix C.

5.3 Data analysis

Several methods were used to analyze the survey data. All the survey data were coded into Microsoft Excel and analyzed in R-studio, a powerful tool for integrating the development environment for the R language. All statistical analyses were tested at least 95% significance.

5.3.1 Descriptive statistics

Descriptive statistics are calculated as a percentage, mean, and standard deviation to reveal the characteristics and preferences of park visitors. Demographic details of visitors, including socioeconomic and education levels, were calculated as percentages. Motivations, benefits, and satisfaction were calculated to compare the factors. From the calculated mean ratings, the order of each factor in motivation, benefits, and satisfaction was identified.

5.3.2 Analysis of variance

Analysis of variance (ANOVA) is often used to find differences among variables. It is an effective way to examine differences between sub-groups and can be used to study differences between linear parks.

5.3.3 Factor analysis

After the descriptive statistics were completed, factor analysis was conducted to identify the reliability of each factor and to combine related ones to test the relationships between factors. Because each factor—such as motivation, benefits, or satisfaction—was asked about via several (6 to 9) variables, these variables can be combined as a single factor to examine causality. Factor analysis is used to reduce the number of variables by grouping them.

In this study, I test each variable's communality to eliminate unrelated sub-questions, and I conduct factor analysis to derive factors by combining variables. The factor analysis proceeds as follows: First, the principal components are calculated using the correlation matrix. If the eigenvalue of the factor analysis is greater than 1.0, it is counted as several factors. Second, varimax is used for the rotation procedure. If the commonality is less than 0.3, the variables are deleted. Third, the factor analysis is conducted again to check the Cronbach alpha value; if it increases after a couple of variables are deleted, the reliability of the model has improved. Finally, grouped factors are derived from the factor analysis.

6. Comparative Study

In phase three, I compare the survey results with social media analytic results by comparing orders between the three parks. This is not a new technique. Wang, Jin, Li, and Zhang (2018) compared social media and survey data to assess the attractiveness of Olympic Forest Park and to study the differences between social media and survey data. Keywords that represent "attractiveness" were extracted from social media data and then asked about in a survey. The authors then compared the order of the keywords between the two cases to study their relationship. As a result, they recommended using social media data first to develop an assessment model and then using a survey to gather supplementary details. These authors used order of keywords to compare two data sets; I do the same here.

6.1 Data collection

This subsection mainly discusses the data collection for phase three. The main data come from the answers to research questions 1 and 2. An answer to question 1 is used to order park visitors' behaviors. Among the components of behavior, only activities are used in phase three because of a lack of information about motivation in social media. Activities identified in question 1 (physical, overlooking, nature, etc.), and the activities of social media users as identified from tweets are collected to compare their the orders. In addition, park visitors' satisfaction with activities and social media users' positive sentiments about activities, collected from surveys and social media analytics, are compared for their order between the three parks.

6.2 Data analysis

The data analysis in phase three follows Wang et al.'s (2018) model. Wang et al. compared social media data with survey data by comparing the order. In phase three, I use behaviors and perceptions to comparison these two methods. Using the numbers of participants in each activity, I order the activities for each park and compare the results. The activities and orderings are detected from social media, and compared to reveal similarities and differences and areas where one method might replace the other.

CHAPTER IV. RESULTS

This chapter describes the results of the study. It is divided into three sections, covering (1) the results of the survey, (2) the results of the social media analytics, (3) a comparison of the two, and (4) the practical implications of this work. The objectives of this study are (1) to determine park visitors' activities and satisfaction using social media data, (2) to compare the results of social media analytics with those of surveys, and (3) to identify the types of data needed by landscape architects according to park type. The first section of this chapter describes findings from the survey. The second addresses park visitors' activities, attitudes, and satisfaction using social media analytics. The third uses order comparison between social media analytics and a survey to determine the validity of social media analytics. The fourth section provides users' information and their purposes in posting their experiences to social media. This section is intended to provide practical results for landscape architects.

1. What Information can Landscape Architects Get from a Survey?

1.1 Who responds to the survey?

The surveys reveal several demographic differences between the three parks (Table 7). Females (54.93%) tended to visit all three parks more than males (45.07%); the High Line had a particularly high proportion of females (56.45%). Age ranges did differ by park, though. At the 606, 33.12% of visitors were 18−29 years old and 26.25% were 30–39. People over 50 visited the High Bridge Trail (36.50%) more than people under 29 (33.12%). And white visitors were more common (68.13%) than other races at all three parks. At High Bridge Trail, white visitors made up 80% of respondents.

The demographics of the respondents differed between the parks, but not significantly.

Visitors to the 606 tended to be white, younger, more educated, and higher-income. Visitors to the High Line, were racially more diverse, moderately educated, and higher-income. High Bridge visitors tended to be white, slightly older, moderately educated, and middle-income.

Table 7. Visitor characteristics in linear parks

	The 606	High Line	High Bridge	Total
Gender				
Male	45.31%	43.55	48.33	45.07
	(n = 141)	(135)	(58)	(334)
Female	54.69	56.45	51.67	54.93
	(175)	(175)	(62)	(412)
Age				
18–29	33.12	24.52	25.83	28.40
	(106)	(76)	(31)	(213)
30–39	26.25	30.65	19.17	26.93
	(84)	(95)	(23)	(202)
40–49	12.19	18.06	17.50	15.47
	(39)	(56)	(21)	(116)
50–59	13.75	12.90	15.83	13.73
	(44)	(40)	(19)	(103)
60–69	10.94	8.71	14.17	10.53
	(35)	(27)	(17)	(79)
Over 70	2.50	5.16	7.50	4.40
	(8)	(16)	(9)	(33)
Prefer not to say	1.25	0.00	0.00	0.53
	(4)	(0)	(0)	(4)
Race				
White	72.50	59.03	80.00	68.13
	(232)	(183)	(96)	(511)
Black	1.56	2.26	3.33	2.13
	(5)	(7)	(4)	(16)
Hispanic	13.12	15.48	2.50	12.40
	(42)	(48)	(3)	(93)
Asian	4.38	11.61	4.17	7.33
	(14)	(27)	(5)	(55)
Other[1]	4.28	8.71	7.50	6.67
	(14)	(27)	(9)	(50)
Prefer not to say	4.06	2.90	2.50	3.33

	(13)	(9)	(3)	(25)
Income				
25,000 or less	8.12	10.00	17.50	11.33
	(26)	(31)	(21)	(85)
25,000−50,000	12.50	17.74	17.50	15.47
	(40)	(55)	(21)	(116)
50,000−75,000	14.69	17.10	13.33	15.47
	(47)	(53)	(16)	(116)
75,000−100,000	16.25	11.29	17.50	14.40
	(52)	(35)	(21)	(108)
100,000−150,000	18.44	20.65	20.83	19.73
	(59)	(64)	(25)	(148)
Over 150,000	19.69	9.35	10.83	14.00
	(63)	(29)	(13)	(105)
Prefer not to say	8.12	13.87	2.50	9.60
	(26)	(43)	(3)	(72)
Education				
High school[2]	10.32	10.32	19.16	13.21
	(33)	(32)	(23)	(88)
University	47.19	57.54	43.33	50.93
	(151)	(179)	(52)	(382)
Graduate school	32.81	23.23	29.17	28.27
	(105)	(72)	(35)	(212)
Postgraduate	5.94	8.71	8.33	7.47
	(19)	(27)	(10)	(56)
Prefer not to say	3.75	0.00	0.00	1.60
	(12)	(0)	(0)	(12)

[1]Responded with nationality or as multiracial.
[2]Including some high school.

One strong points of the survey is that it provides details, such as respondents' gender, age, race, income, and education, that are not available through social media. If planners or designers need to know who is using a park, surveys are a tool for gathering this information.

1.2 What park visitor activities can be investigated using a survey?

Activities in the three linear parks fall into four categories: (1) physical activities,

(2) social activities, (3) relaxation and restoration, and (4) viewing (Table 8). The main activities differed between the park, but overall the greatest number of visitors engaged in physical activities.

Table8. Main activities in each linear park

	The 606	High Line	High Bridge	Total
Physical activity				
Biking	48.13%	4.19	25.83	26.40
Hiking or walking	80.94	76.77	65.83	76.80
Jogging or running	28.75	10.97	5.83	17.73
Walking dogs	17.81	7.74	7.50	12.00
Social activity				
Picnicking	5.94	11.29	0.00	7.20
Posting on social media	9.38	20.65	7.50	13.73
Relaxation and restoration				
Relaxation or no main activity	53.75	63.87	28.33	53.87
Photography	25.31	49.03	70.83	42.4
Viewing				
Viewing from overlooks	21.56	70.83	62.26	46.27
Viewing public art	27.19	40.97	0.00	28.53
Viewing nature	17.50	13.87	49.17	21.07

Overall, the largest portion of visitors to the 606 took part in physical activities. Hiking and walking were the main physical activity (80.94%), followed by cycling (48.13%), and jogging or running (28.75%). High Bridge visitors also engaged in physical activities, such as cycling (25.83%), hiking, and walking (25.83%). But they jogged and ran far less (5.83%). Most visitors to the High Line walked (76.77%); very few rode bikes (4.19%), as biking and dog walking were prohibited on the High Line (Figure 6).

Social activities show the differences between the parks (Figure *7*). For example, visitors to the High Line arrived to have picnics offline (11.29%) and then posted their experiences on social media to share them with others online (20.65%). However, no visitors to the High Bridge had picnics, and few posted about their visits on social media (7.50%). Visitors to the 606 had picnics (5.94%) and posted about the park on social media, but not very actively (9.38%).

People engaging in relaxation, restoration, or no major activities also differed between the parks (Figure *7*). Many people visited the 606 and the High Line to relax (53.75%, 63.87%, respectively), but visitors to the High Bridge did so less (28.33%). Photography as a recreational activity also differed among the parks. Only 25.31% visitors to the 606 took photos, far fewer than in the other parks; 70.83% of High Bridge visitors took photos.

Viewing activities were measured in terms of three themes (Figure 8): viewing from overlooks or observation points, enjoying public art, and viewing nature. Visitors to the High Line enjoyed observatory points (70.83%) and public arts (40.97%) more than those at the other parks. On the High Bridge Trail, many visitors viewed from overlooks (62.26%) and enjoyed natural scenery (49.17%), but none viewed public art. Visitors to the 606 moderately enjoyed overlooks (21.56%), public arts (27.19%), and nature (17.5%).

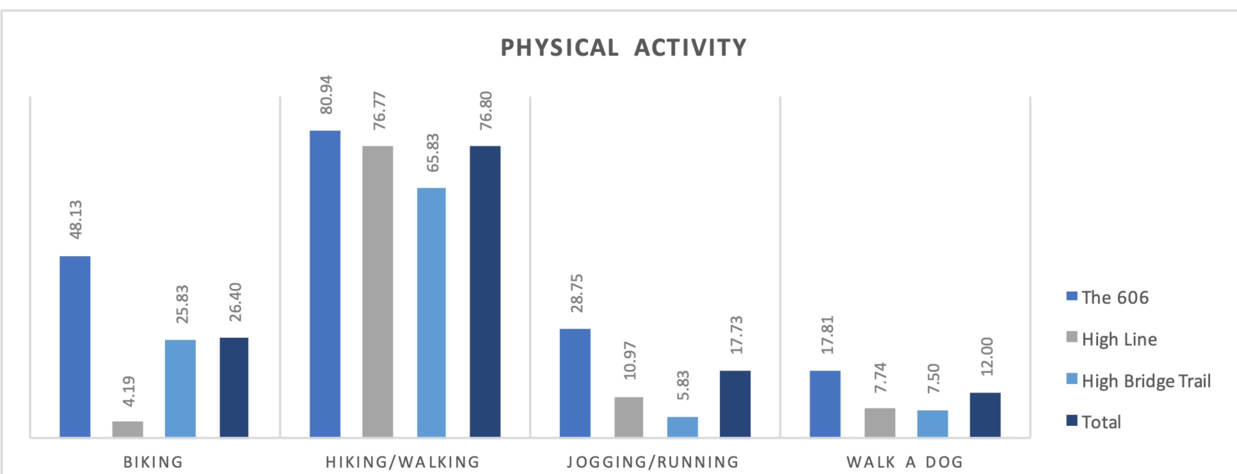


Figure 6. *Physical activities.*

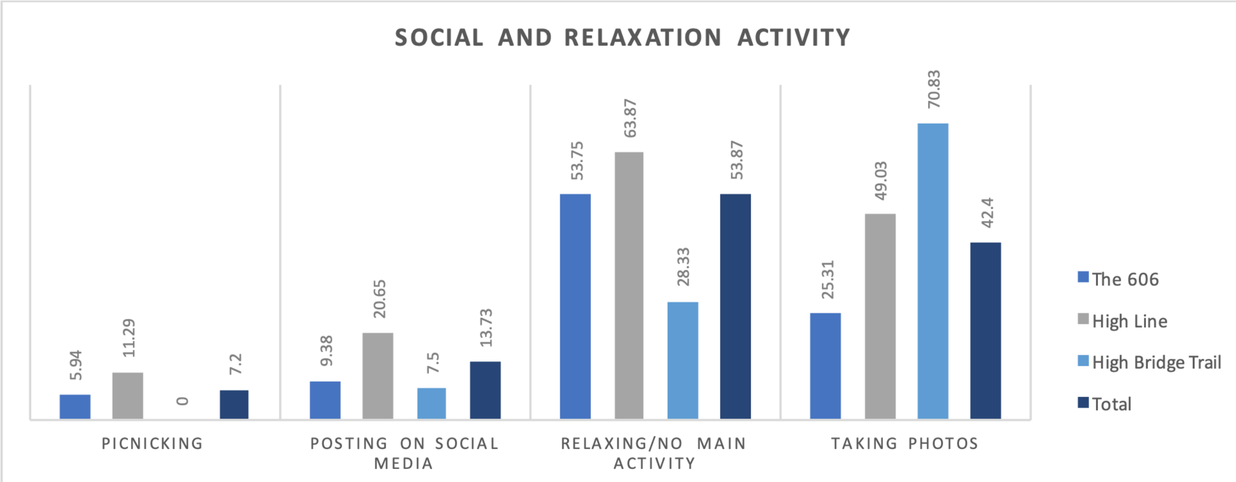


Figure 7. Social and relaxation activities.

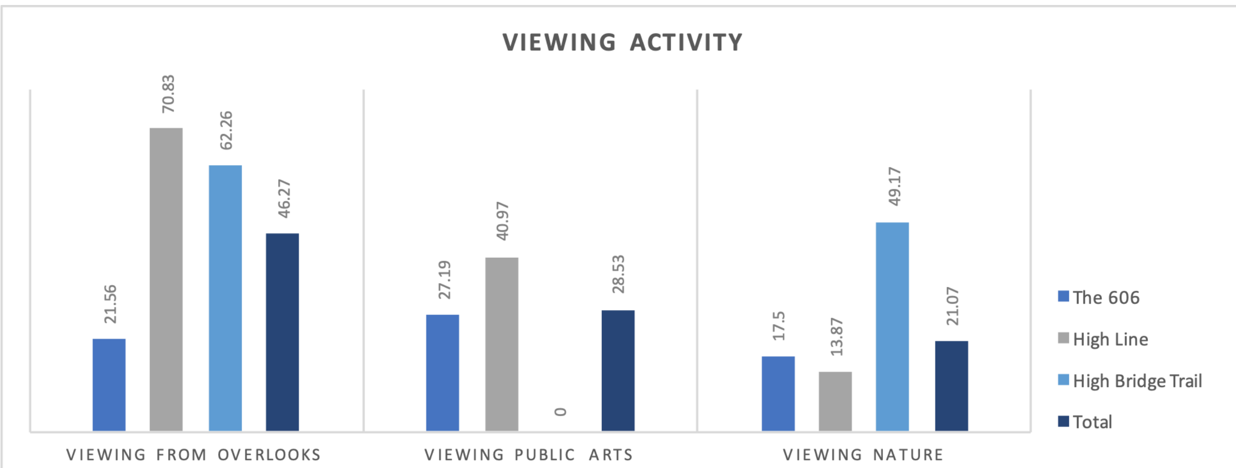


Figure 8. Viewing activities.

In summary, visitors to the 606 usually participated in physical activities, especially biking and walking, and rarely in social activities. They liked to relax and take photos in the park. They also enjoyed viewing from overlooks, seeing public art, and encountering nature, but less than at the other parks. Visitors to the 606 were most likely overall to engage in things that could be done individually, such as biking, walking, relaxing and viewing the scenery.

At the High Line, most visitors also engaged in physical activity, but the main activities were hiking, walking, jogging, and running. People visited mainly to relax and see interesting things, such as public art or urban scenery from observatory points. Visitors also engaged in more social activity than at other parks. Overall, visitors of the High Line came to the park mainly (1) to rest, (2) to walk, run, or jog, (3) to have picnics, (4) to interact with others through social media, and (5) to enjoy public arts and urban scenery.

On the High Bridge Trail, similarly, physical activity was very popular. Visitors rarely picnicked or posted on social media about their experiences, however. Visitors also enjoyed viewing nature and watching from overlooks, but not public art. Overall, visitors engaged in physical activity or took photos at overlook points and enjoyed the scenery.

Surveys can be identified by allowing park visitors to double check what they do in the park. This method is widely used and has the advantage of identifying people in the park. However, a limitation is that the surveyor must to know in advance what people are doing so that accurate investigations are possible. For instance, behavior such as skateboarding on the 606 cannot be identified unless asked about in advance on the survey. Thus what people do in a park can be determined using a prepared survey, but a drawback is not knowing about things that aren't in the responses.

1.3 What park visitor benefits can be investigated using a survey?

Through PCA, six variables were combined into three factors (Table 9): natural benefits, social benefits, and health benefits. The natural benefits of a park help visitors understand nature better (Q10.1) and enjoy natural scenery (Q10.2). Social benefits help visitors interact with others (Q10.3) and build or strengthen relationship (Q10.4). Health benefits are elements that improve physical (Q10.5) and mental health (Q10.6). Descriptive statistics were calculated for these three factors (Table 10 and Figure 9).

Visitors to the High Bridge Trail perceived natural benefits (mean = 4.17). Visitors to the

High Line perceived all three benefits less than at the other parks. Visitors to the 606 perceived social (3.53) and health benefits (4.51), more than at the other parks. The differences between all the parks and factors were statistically significant (Table 11).

Table9. Results of the principal component analysis of benefits

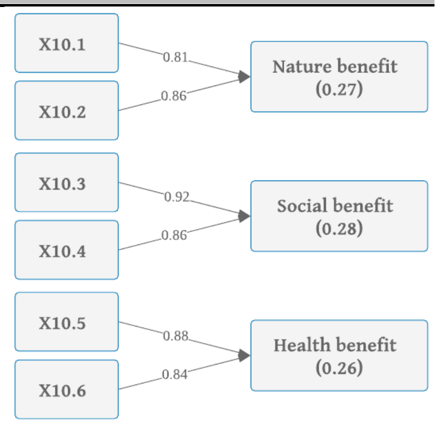
	m	Median	Communality	Cronbach alpha (0.78)	Loading and selected factors
Q10.1	3.23	3.00	0.76	0.74	
Q10.2	4.23	4.00	0.81	0.75	
Q10.3	3.28	3.00	0.86	0.77	
Q10.4	3.33	3.00	0.82	0.75	
Q10.5	4.07	4.00	0.84	0.75	
Q10.6	4.23	4.00	0.82	0.75	

Table 10. Descriptive statistics of benefits

	Count	Mean	Median	Sd.
Natural benefit				
The 606	320	3.89	4.0	0.74
High Line	310	3.42	4.0	0.51
High Bridge Trail	120	4.17	3.5	0.89
Total	750	3.73	4.0	0.83
Social benefit				
The 606	320	3.53	3.5	0.86
High Line	310	3.06	3.0	0.89
High Bridge Trail	120	3.31	3.5	1.29
Total	750	3.30	3.0	0.98
Health benefit				
The 606	320	4.51	5.0	0.58
High Line	310	3.66	4.0	0.85
High Bridge Trail	120	4.47	4.5	0.91
Total	750	4.15	4.0	0.83

Nature benefit **Social benefit** **Health benefit**

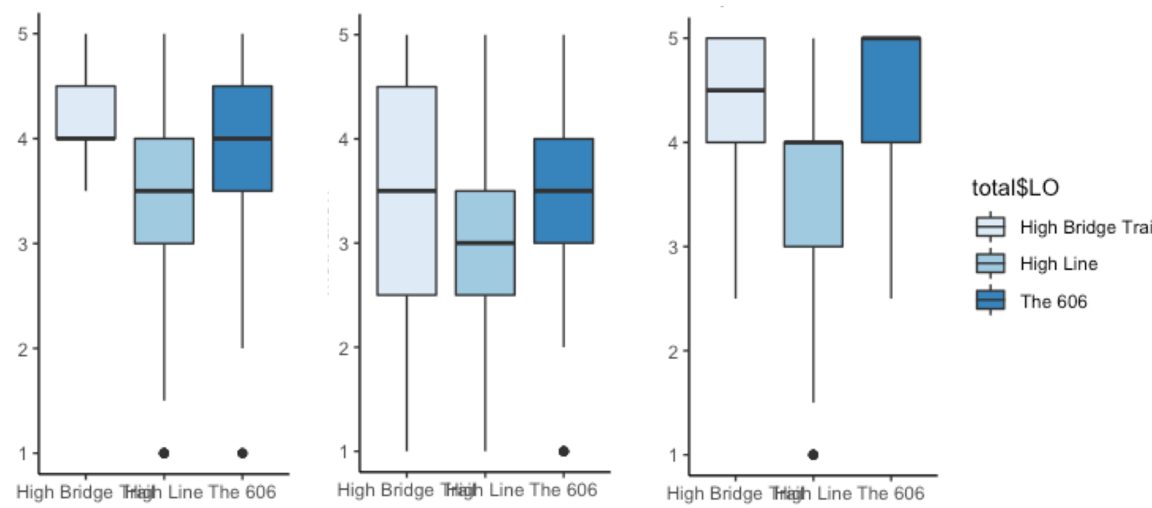


Figure 9. Box plots of benefits.

Table 11. Differences between the three parks (results of ANOVA)

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
Nature benefit	2	59.5	29.75	48.5	<2e-16 ***
Residuals	747	458.2	0.613		
Social benefit	2	35.4	17.696	19.39	6.17e-09 ***
Residuals	744	678.9	0.912		
Health benefit	2	127.5	63.77	122	<2e-16 ***
Residuals	747	390.4	0.52		

0 ‘***’ 0.001 ‘**’ 0.01 ‘*’

The advantage of the survey is that it can be analyzed statistically. As in this study, researchers can compare park numbers on the same scale and analyze the differences. For example, this study asked park visitors about benefits and measured these on a 5-point scale. This 5-point scale is suitable for comparing responses between parks.

2. What information can landscape architects get from social media?

This question investigates information that can be gleaned from textual data in social media for use in landscape architecture. This information has three main components: (1) activities, (2) interests, and (3) satisfaction.

Regarding the 606, the number of tweets and their main topics reflect the park's location in the midst of residential areas. The five most-tweeted dates show the main issues people are interested in. Two of the abnormal dates were events day for residents: the first anniversary of the 606, and the announcement of a block party on the 606. The topics on two others days were related to crimes (thefts) on the 606. One date marked a protest against rising rents, property prices, and taxes. Users also mentioned the 606 the most in May, July, September, and October, and the overall number of tweets has decreased each year. The 606 opened in June 2015, and the collected tweets run from August 2015 to July 2019 (Figure 10).

The High Line has the largest textual data corpus (n = 165,347) of the three, almost eleven times as many as the 606 (n = 14,340). Abnormal dates were again reviewed to identify major events or issues in the park, and there was only one among the top ten dates in number of tweets: the fifth annual Coach and Friends of the High Line summer party on July 25, 2015. The other dates are also from 2015, which is the top-tweeted year of the study period (Figure 11).

The High Bridge trail has only 457 tweets from 2015–2019, the fewest of the three. Because of the low number, abnormal dates and heavy users were not identified. The numbers of tweets by month and year also show a different trend from the other two parks. The 606 and the High Line do not show much difference between months, but the High Bridge shows big differences between summer and winter. By year, High Bridge tweets decrease from 2015 to 2017, but they increase in 2018. As in the other linear parks, positive sentiment increases from 2016 onward. An interesting point occurs in 2016: for some reason, the positive tweets decrease dramatically from 2015 to 2016 (Figure 12).

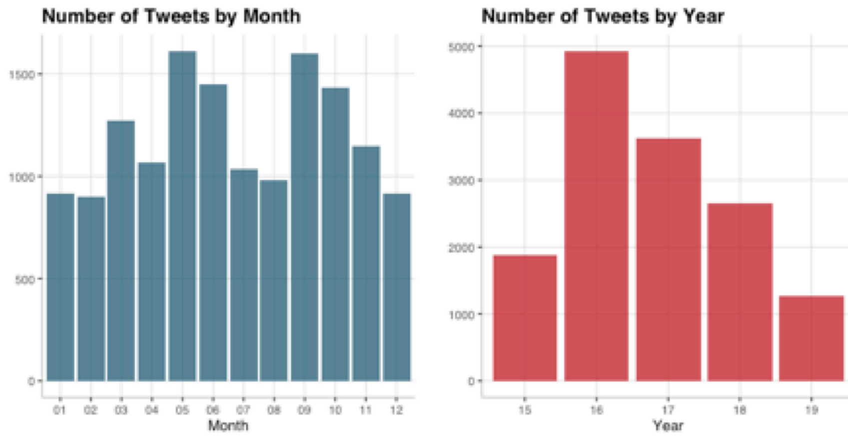


Figure 10. The 606: (a) number of tweets by month, (b) number of tweets by year.

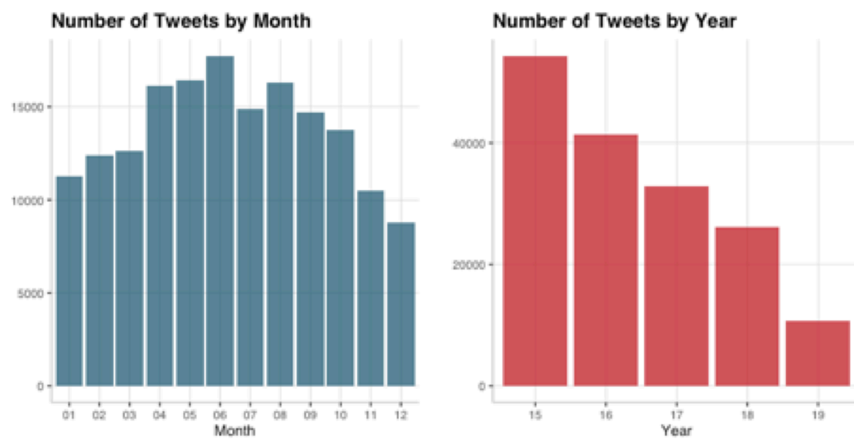


Figure 11.The High Line: (a) number of tweets by month, (b) number of tweets by year.

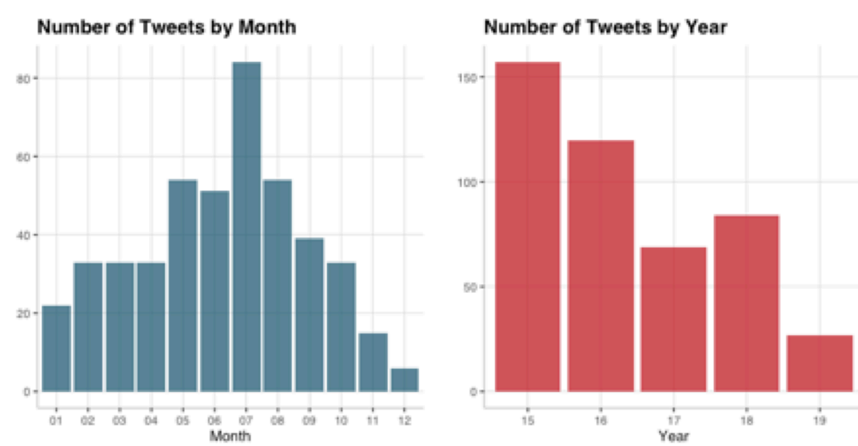


Figure 12. The High Bridge: (a) number of tweets of by month, (b) by year.

2.1 What park visitor activities can be investigated using social media analytics?

This section investigates social media users' activities in the parks using keyword searches. According to the survey, the main activities in the parks physical activity, art viewing, social interaction, picnics, viewing from overlooks, and encountering nature (Table 12). Several activity-related keywords were carefully selected using the surveys. For example, "walk," "jog," "run," and "bike" were selected because those words were used in the survey questions. The selected keywords were used to filter activity-related tweets. Table 12 shows the selected keywords and numbers of tweets mentioning them.

Table 12. Number of tweets related to the activity

Activity	keywords	Number of tweets		
		High Line	The 606	High Bridge
Physical activity	"walk," "jog," "run," "bike"	10509 (6.36%)	730 (5.64)	42 (9.19)
Art	"art," "gallery," "museum"	15021 (9.08)	987 (7.62)	18 (3.94)
Social	"friend," "family," "mom," "dad," "people," "father," "mother"	2408 (1.46)	177 (1.37)	6 (1.31)
Picnic	"lunch," "snack," "sandwich," "food," "coffee," "yum"	445 (0.27)	45 (0.35)	7 (1.97)
Overlooking	"overlook," "view," "see," "watch"	275 (0.17)	4 (0.03)	0 (0.00)
Nature	"trees," "wild," "planting," "garden," "vegetation," "nature"	4005 (2.42)	176 (1.36)	3 (0.66)

One strength of big data is that it allows researchers to trace past trends. Analyzing activities by month and year in the High Line revealed several things. First, art viewing and social activities on the High Line have several similarities. For example, both have been increasing since 2017. These similarities also can be found in the activity graph by month (Figure 13). This can be interpreted as meaning that social media users tend to visit the park and enjoy public art with others.

Second, physical activity has been increasing rapidly since 2017. This can be seen in Figure 5: the physical activity percentage in winter is higher than in summer, which can be interpreted two ways. Social media users may post about their physical activities more in winter, or they make engage in more physical activity in winter, perhaps because of crowds in summer.

Third, picnic and viewing activities, especially viewing from overlooks, are the least popular activities. Social media users rarely mentioned overlook views or picnics. This too can be interpreted in two ways. Social media may rarely take part in those activities, or they may take part but not post about them. Although High Line users had picnics and enjoyed overlooks comparably to users of the other two parks, these activities were less popular than any others.

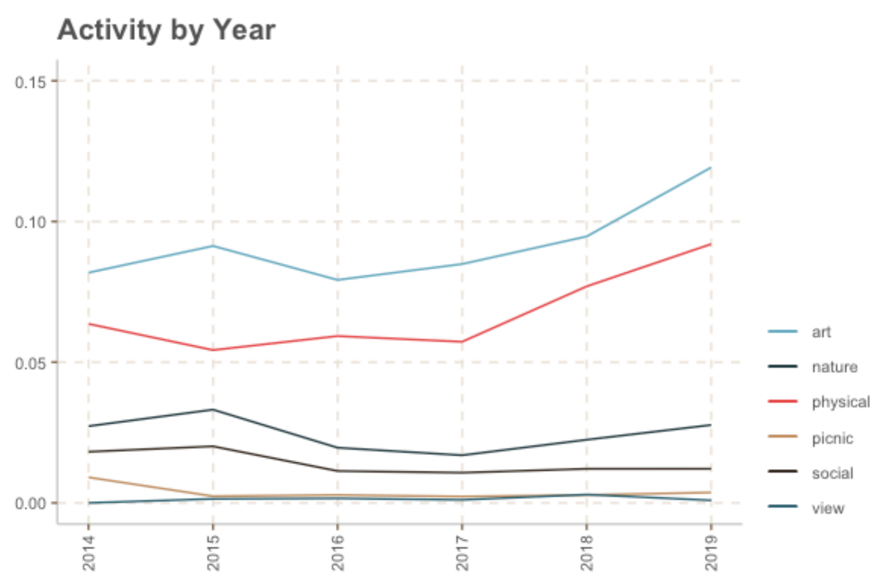


Figure 13. The High Line: Activities by year (x: years, y: activity/total tweets).

On the 606, most activities increased from 2016 to 2018, but have been decreasing since. Art viewing in particular decreased dramatically since 2018, even though the 606 has launched public art programs. Physical activity has increased slightly since 2016. One interesting thing is the relationship between “arts” and “nature.” These two kinds of activities show similar patterns (Figure 14). This may mean that people who visit linear parks to look at art or nature tend to see the other too.

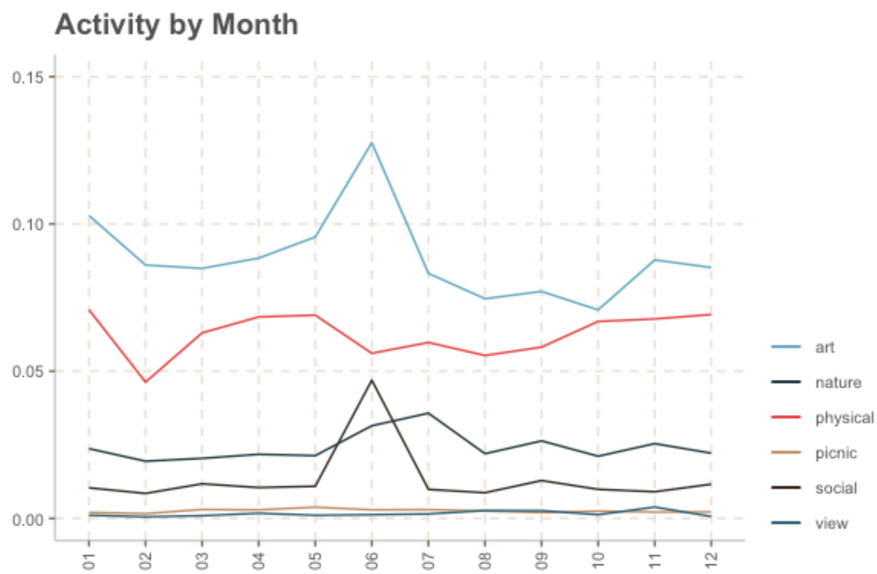


Figure 14. The High Line: Activities by month (x: months, y: activity/total tweets).

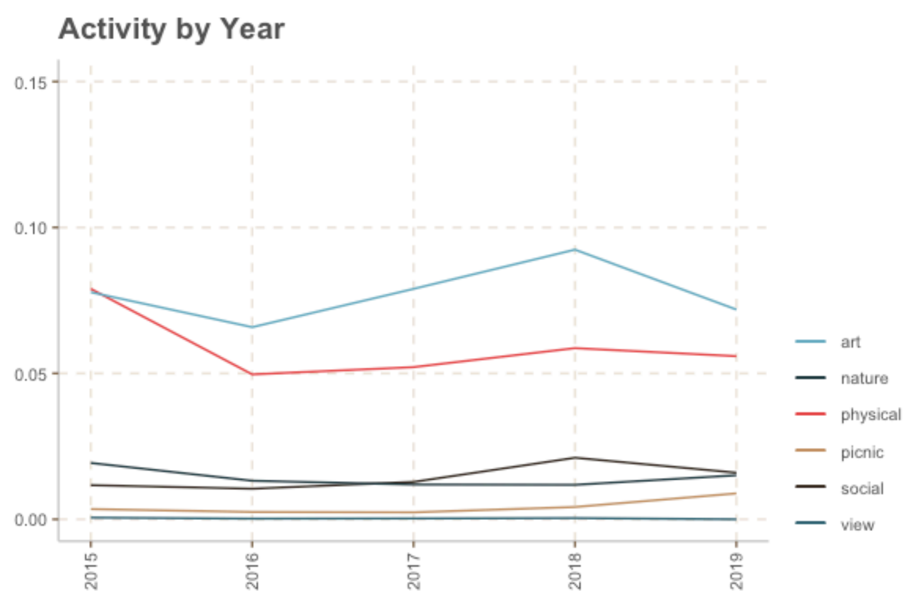


Figure 15. The 606: Activities by year (x: years, y: activity/total tweets).

Activities by month also show some interest features (Figure *16*). Art viewing was concentrated in the summer and physical activity was high in the winter. This can be interpreted as meaning that art viewing is not related to physical activity on the 606.

A benefit of social media analytics is in letting us view historical data and identify trends. As these results show, social media analysis reveals changes in behavior by year and month. By contrast, surveys are limited in not providing historical data, because they only carry data from the time they are completed. This advantage could make social media useful for managing parks.

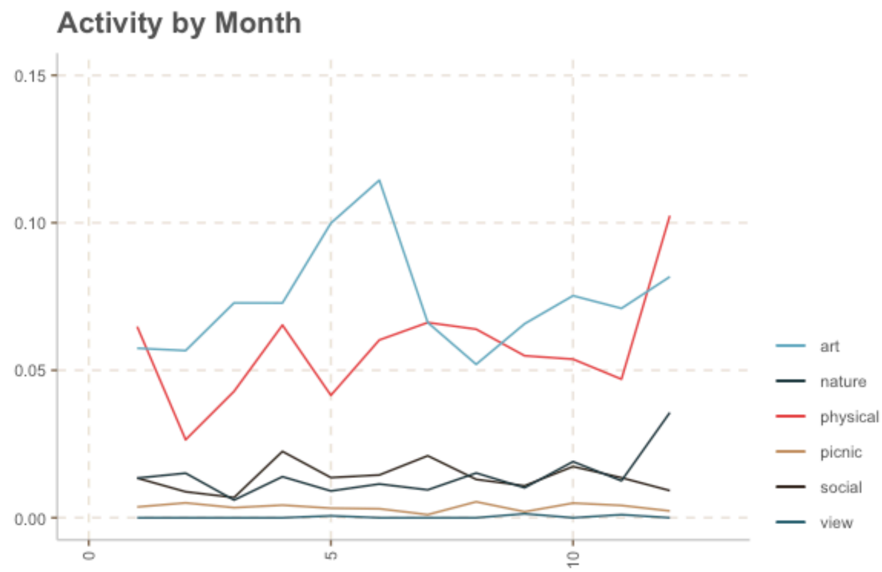


Figure 16. *The 606: Activities by month (x: months, y: activity/total tweets).*

2.2 What park visitor interests can be investigated using social media analytics?

This question focuses on public opinion about linear parks, which is valuable for tracing the meaning of a place to people. Because social media platforms let users share their daily lives, their data can be valuable source of representing what users do in a day. Using social media analytics, this section describes the main issues in each linear park by topic analysis.

In topic changes by month, the High Line is more monotonous than the 606. Figure 17 shows the results of the topic analysis. People using the High Line mentioned “art,” “walk,” “photo,” “people,” and “city” frequently. It can be presumed that one reasons for the park’s success is New York City itself; art could be another reason. High Line users like taking photos in all seasons, and people mentioned “photo” frequently. This means the High Line is a place to memorialize by posting on social media, not an everyday place. The urban views on the High Line may also be loved; High Line users mentioned “building” and “tower” frequently.

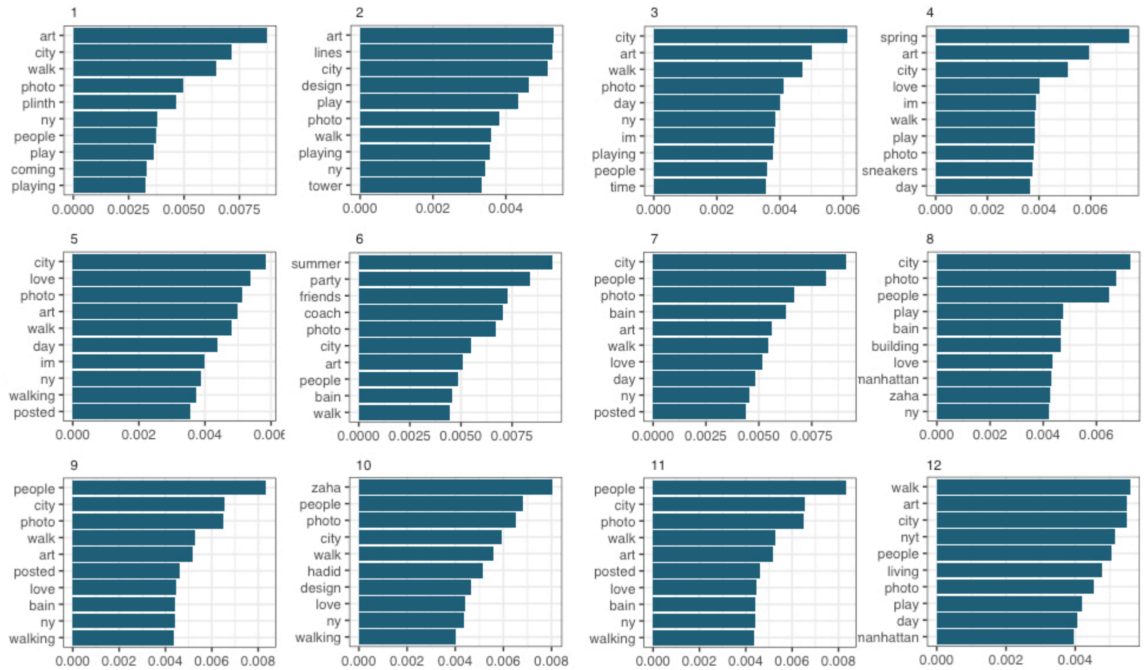


Figure 17. Topic analysis of the High Line by month (x: topic/total tweets, y: topic).

The topic analysis presents changes in topic by month (Figure 18). In the winter, from December to February, park users frequently mention “light” and “walk” (or “walking”), which could mean the 606 is used for light walks in the winter. “Jazz” is also mentioned more in January and February than in other seasons. The main topics of the spring appear to be worries: people tweeted about “robberies,” “crime,” “gentrification,” “police,” “night,” and “housing” from March to May. They also mentioned “tensions” and “buzzing” to express their feelings. These findings may reflect 606 users worrying about housing prices and crimes near the park. Also in the spring, April has different results from the other months: people mention activities, such as “bike” and “walk,” and their interests: “free,” “jazz,” and “night.” In summer, park users mention physical activities, such as “walk,” “bike,” and “run,” along with their status, such as “happiness.” In the fall, people liked to take “photos” more than in other seasons. They also tweet “beautiful” in November, which is the first month they describe feelings about the scenery.

Overall, users of the 606 park worry about “housing,” “prices,” and “crimes.” They like to do physical activities in all seasons, though the intensity may differ, from light walking in the winter to running in the summer. Lastly, people enjoy taking photos in the fall.

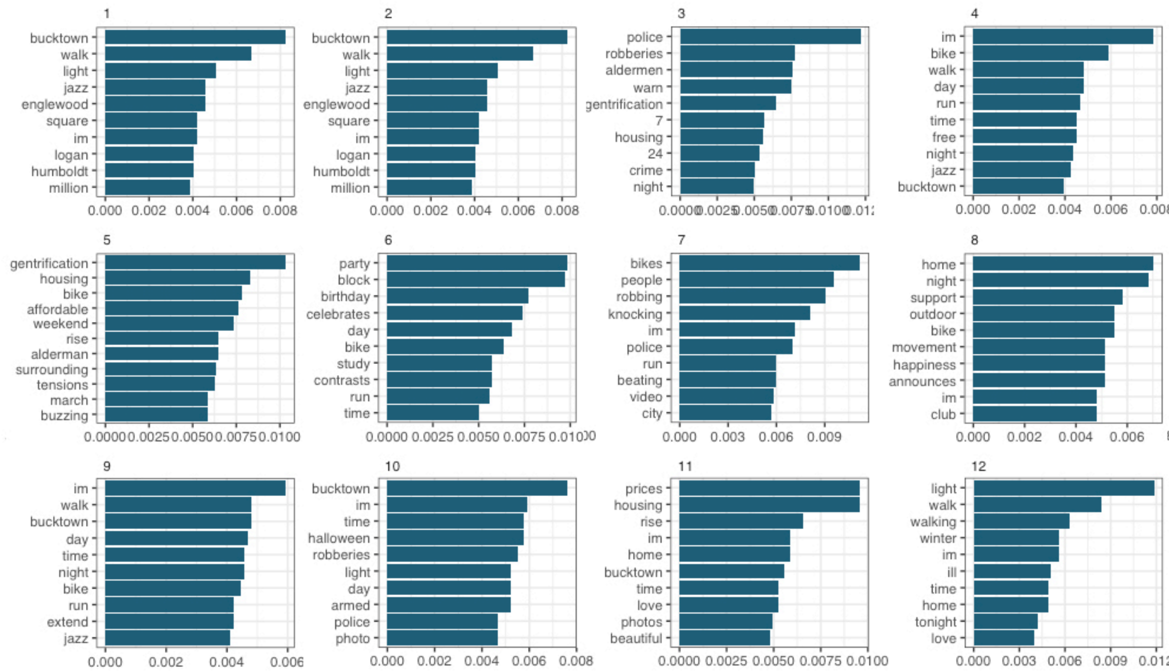


Figure 18. *Topic analysis of the 606 by month.*

Topic analysis is a good way to find out what people are talking about. It is a way to grasp not the economic effects of the park, but the way it exists in people’s daily lives, by collecting and analyzing the information that users talk about. This type of analysis is possible through social media analysis but not through surveys: questionnaires must be formulated in advance, and only the answers to the chosen questions are collected.

2.3 What park visitor benefits can be investigated using social media analytics?

This section describes the main issues in each linear park by users’ sentiments. First, trends and sentiments are represented to trace how frequently social media users mention linear parks (*number of tweets*) and what they think about them (*sentiment*). We compare sentiments by month and day to identify trends in how people feel. Second, we calculated term frequency to identify the main topics of tweets as a medium for understanding public opinion. We also divide the tweets into polarity sentiments to identify the main theme of each sentiment, negative or positive. These let us identify the causes of negative and positive experiences.

Users tend to mention the 606 positively. The histogram in Figure 19 shows positive sentiments increasing annually, although the total number of tweets decreases over the years. One interesting thing is the difference between summer and winter. Sentiment differ only slightly by month, but a higher percentage of tweets are positive in July and December than in other months. Because winter in Chicago is severely cold, this is an unexpected result.

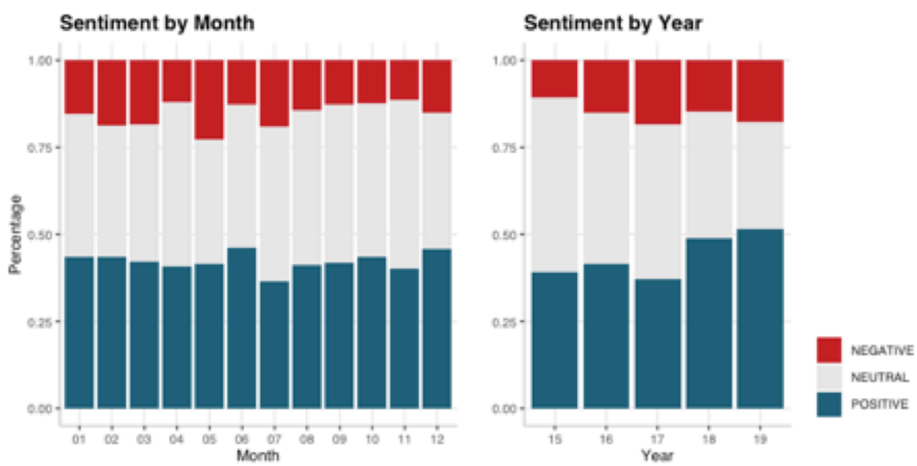


Figure 19. *The 606: (a) sentiment by month, (b) sentiment by year.*

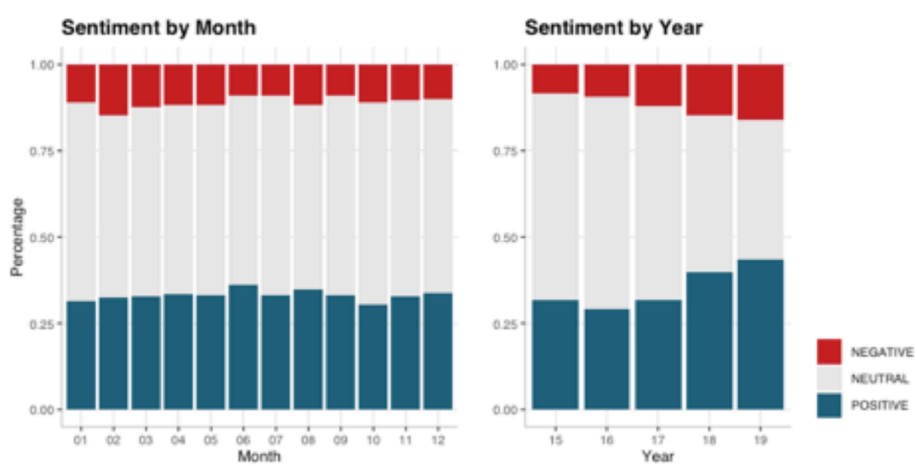


Figure 20. *The High Line: (a) sentiment by month, (b) sentiment by year.*

High Line users tend to post their experiences on social media platforms more in the summer and fall than 606 users do; however, the annual number of tweets declines from 2015 to 2019 (Figure 20). Users also tend to mention the High Line more positively each year. But month by month, there are only small changes in sentiment.

At High Bridge, positive sentiments increase from 2016 onward (Figure 21). Interestingly, social media users tend to post about the High Bridge less positively in summer than in other seasons.

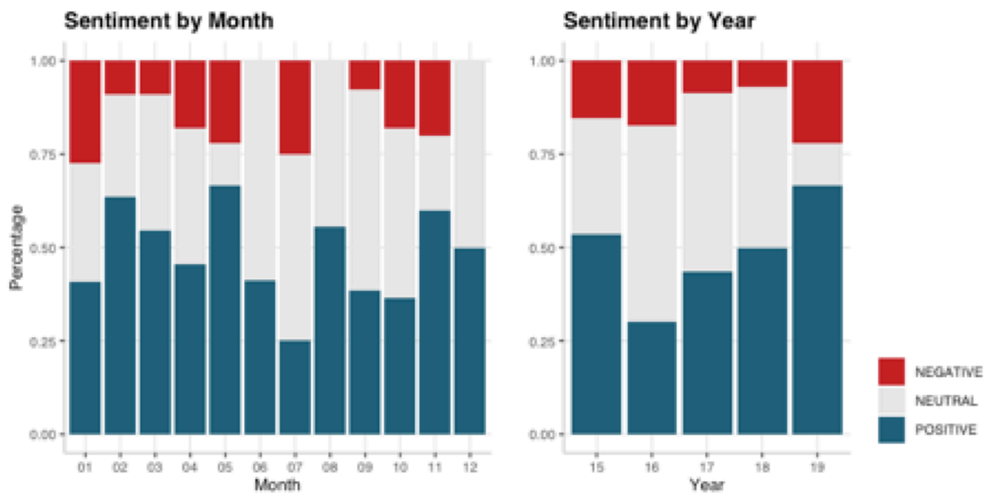


Figure 21. The High Bridge: (a) sentiments by month, (b) sentiments by year.

Monthly and yearly sentiment analyses are a good indicator of how park visitors change. In particular, they are helpful for judging how satisfaction changes monthly and yearly. It is expensive to conduct the same survey monthly and yearly, but social media analytics can include historical data an no extra cost. The analyzed information can be used to gauge user satisfaction or the effectiveness of specific park programs, which makes social media data suitable for emotional analysis.

I also conduct a bigram network analysis to scrutinize tweet sentiments. Users posted positively when the 606 celebrated its first birthday, and for free or cheap events and block parties (Figure 22). Users mention the 606 negatively in reference to protests, gentrification, affordable housing, robberies and the police, and tensions (Figure 23). Those tweets show the issues users worry about. There are mainly two: gentrification and crime.

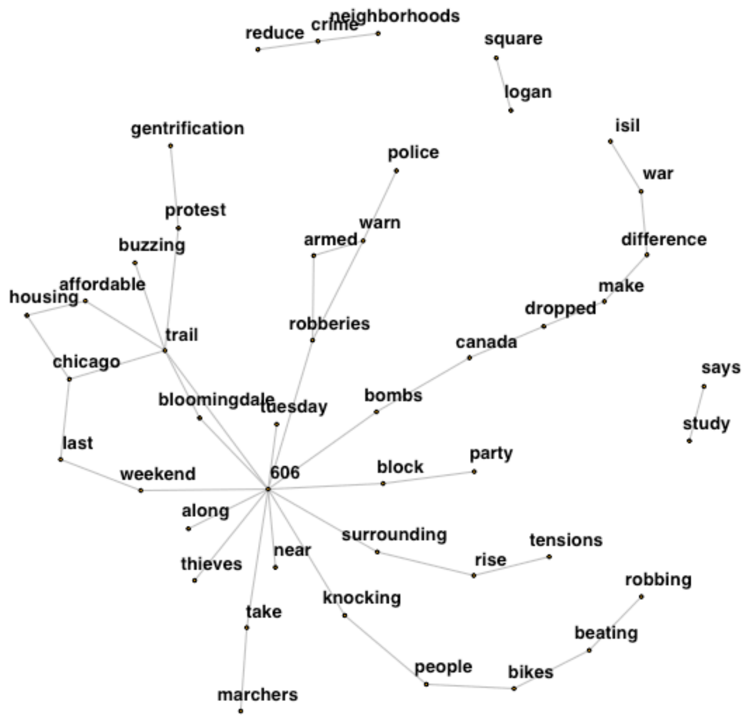


Figure 22. The 606: Bigram network of negative-sentiment tweets.

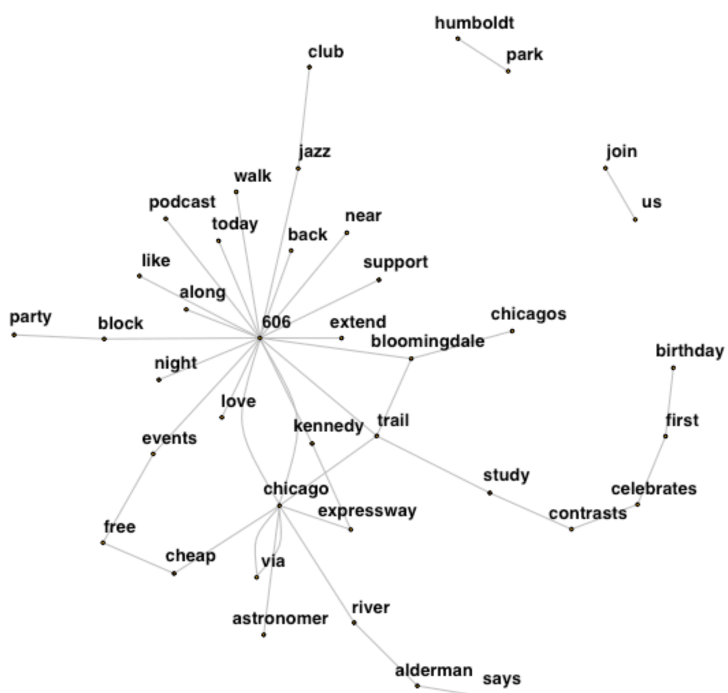


Figure 23. The 606: Bigram network of positive-sentiment tweets.

The bigram network of positive sentiments represents keywords related to the High Line (Figure 24). For example, one cluster includes “im,” “pretty,” and “good.” Some clusters show events on the High Line, including, “friends,” “coach,” “annual,” and “5th.” Interesting tweets with positive sentiments are “serene,” “scenic,” and “bridge.” This cluster suggests that the main point of the High Line is the scenic view of New York City from the bridge. “Central” and “Park” are also frequently mentioned on the High Line. From these results, one of the biggest reasons for the High Line’s success can be presumed to be the scenic urban view and the city itself. As for negative sentiments (Figure 25), the bigram network analysis shows keywords liked “unused” with “railroad” and “Chelsea” with “site.” Some keywords also appear related to the High Line itself, such as “ridiculous” and “suicidal.”

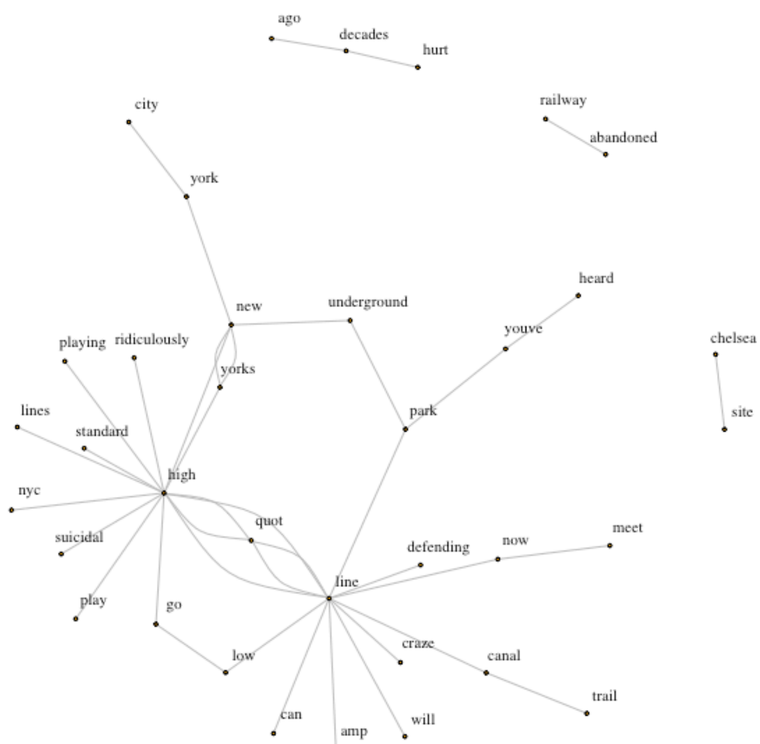


Figure 24. The High Line: Bigram network of negative-sentiment tweets.

Figure 25. The High Line: Bigram network of positive-sentiment tweets.

In the case of the High Bridge, positive sentiments (Figure 26) include mentions of the bridge as "awesome," "beautiful," and "great," along with descriptions of what users did there, such as "run," "riding," and "biking." There were no special issues or event in the park. Users also mentioned the parks background with the terms "railroad," "donated," and "unused"; the last could express negative sentiment (Figure 27).

This section addresses the relationship between users' activities and their attitudes by filtering tweets for specific activities and comparing the sentiments. The main activities in linear parks are physical activities, art viewing, social interaction, picnics, viewing from overlooks, and encountering nature (Table 13). Keywords related to each category are used to filter tweets, and sentiment analysis is used to compare activity-related and unrelated tweets. Because the data volume for the High Bridge is lower than for the other two parks, its results are described separately, but the other two are compared.

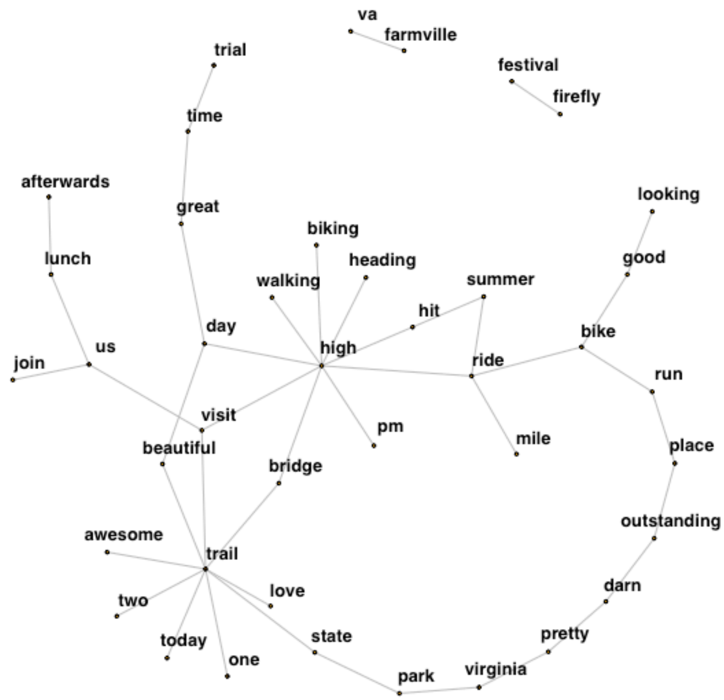


Figure 26. The High Bridge: Bigram network of positive-sentiment tweets.

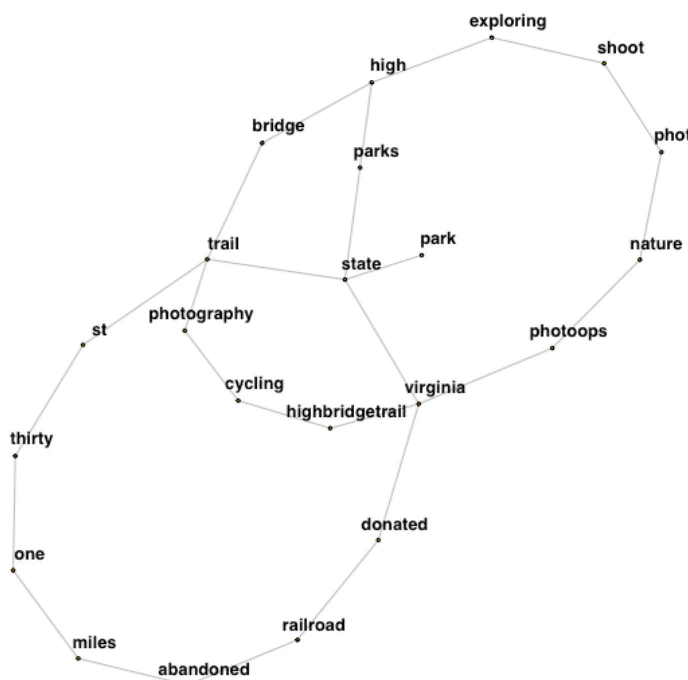
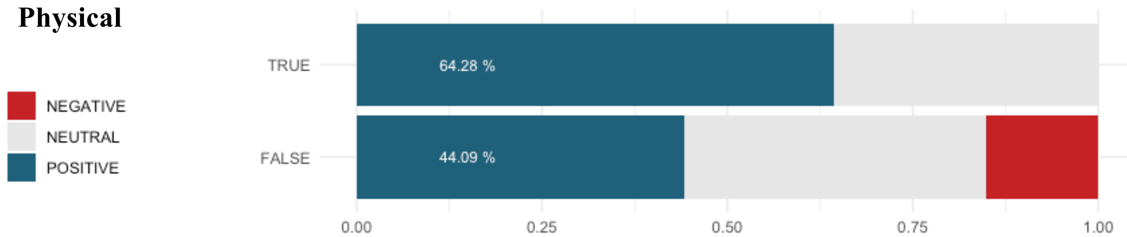
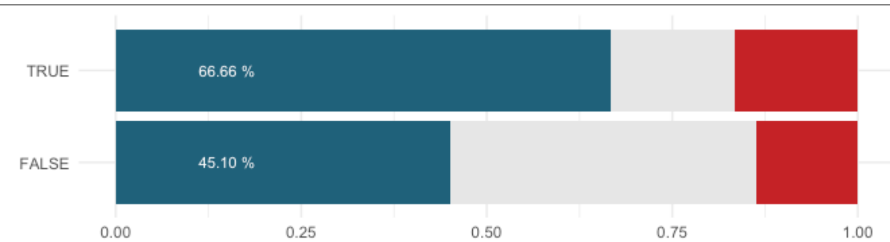
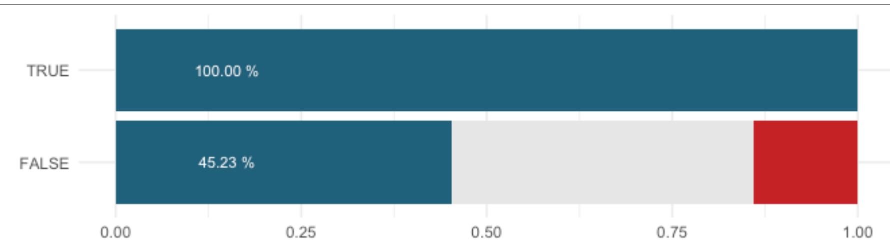
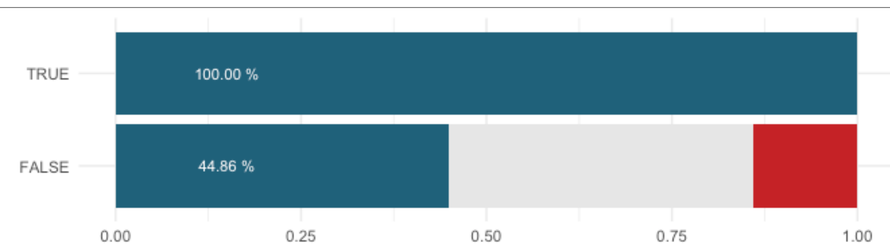


Figure 27. The High Bridge: Bigram network of negative-sentiment tweets.

The results are as follows: (1) High Bridge users posted positively about most activities except viewing from overlooks, (2) 606 users posted positively about physical activities, enjoying art, social interaction, and enjoying nature, and (3) High Line users posted positively about picnics and overlooks. These results suggest that the 606 supports residents using the trail for daily activities, whereas the High Line is a park to eat lunch in and have a special experience in the center of the city with a pleasant urban view.

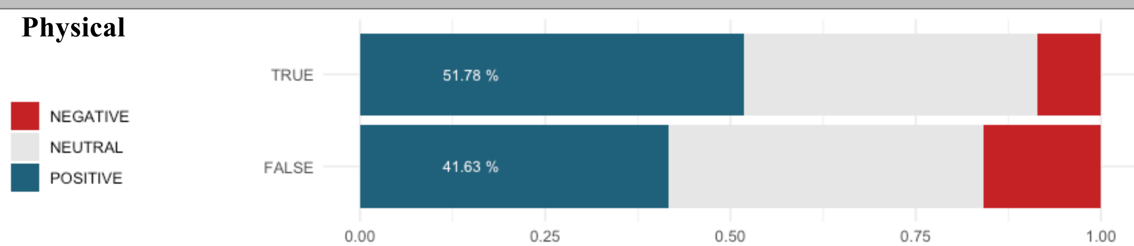
Table 13. Sentiments by activity on the High Bridge

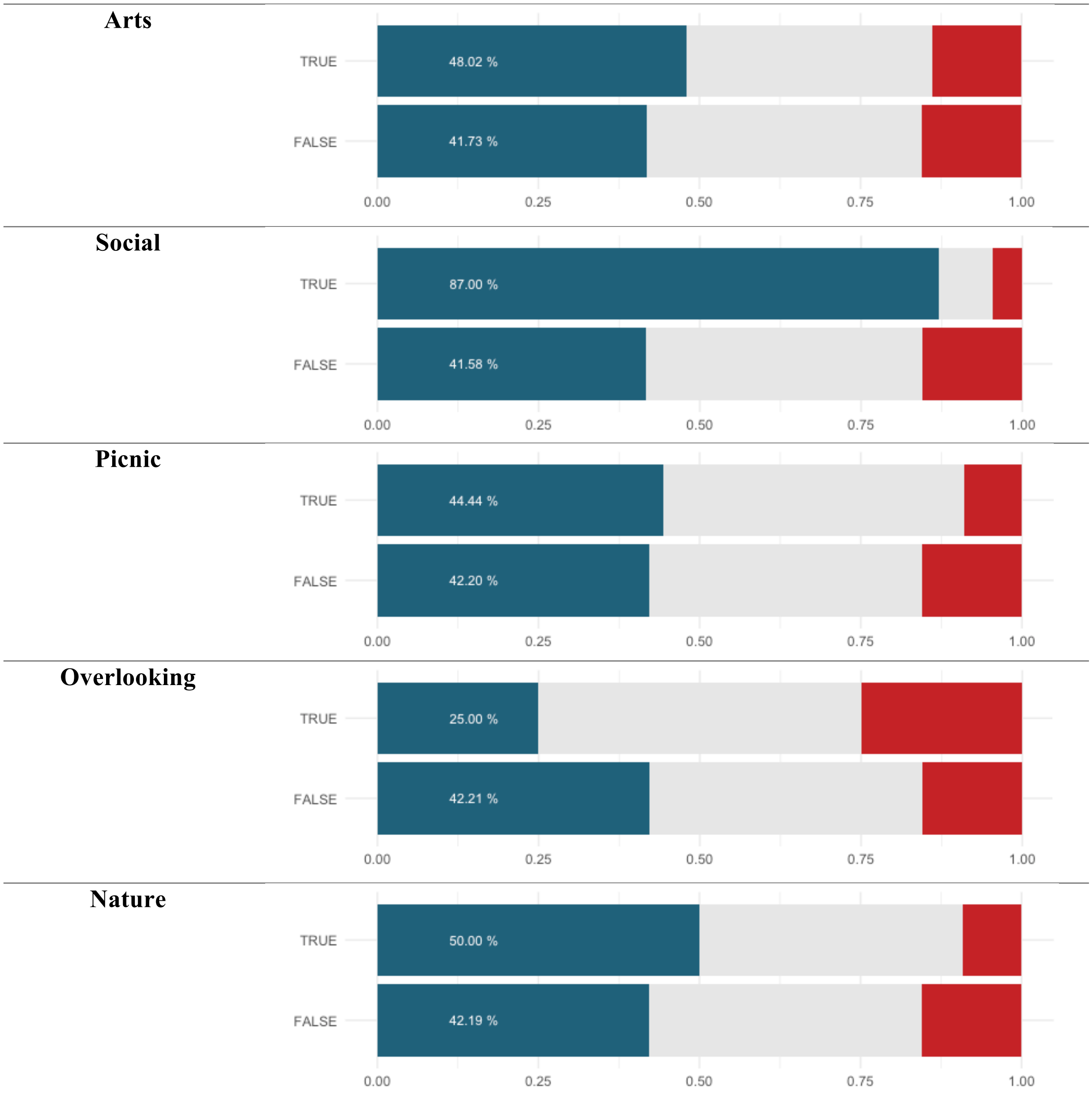
Activity	Sentiments for each activity (True: Doing activity; False: Not doing)
Physical	
Arts	
Social	
Picnic	

First, High Bridge users post positively about all activities except overlooking (Table 28). Because of the physical attributes of the High Bridge, which the highest railway and historical monument in Virginia, its overlooks are a unique characteristic. However, users do not mention the overlook views on social media. Because the High Bridge has fewer tweets than the other two parks, its results tend to be skewed to either the negative or the positive. For example, there are nine tweets about picnics, and all are positive; there are six tweets about social interaction, and all these are positive too. The insufficient data may drive these skewed results.

Users of the 606, engage positively in physical activities, art and nature activities, and social interaction (Table 14). They are also more satisfied with their physical activities (51.78%) than High Line users (30.31%), and are more involved in physical activities overall (24.73%) than High Line users (6.36%). Users of the 606 users also mention the arts (33.43%) more than High Line users (9.08%) and more positively (48.02% vs. 39.74%). Users of both parks tweet positively about social interaction, which includes experiences with friends and family; 606 users slightly more so than High Line users (87.00% vs. 84.50%). Users of the 606 users post more positively (50.00%) than High Line users (40.58%) about interactions with nature. However, the difference in positive sentiment between nature-related tweets and others is similar in the two, around 8%. Overall, 606 users post positively about physical, art, social, and nature activities.

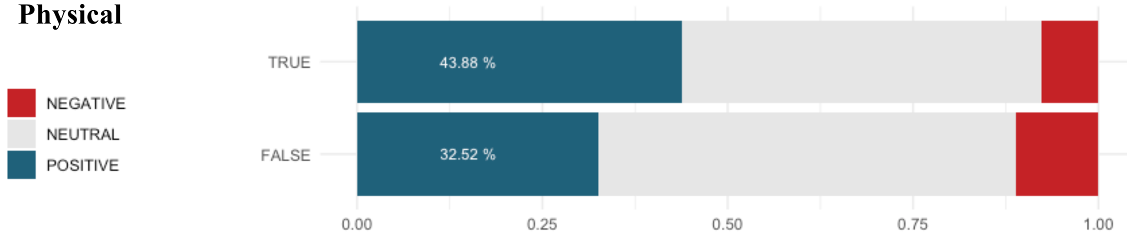
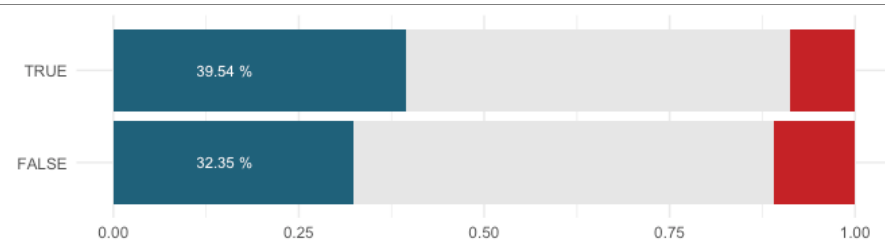
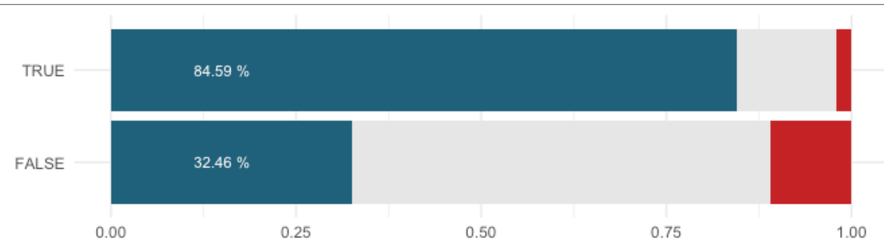
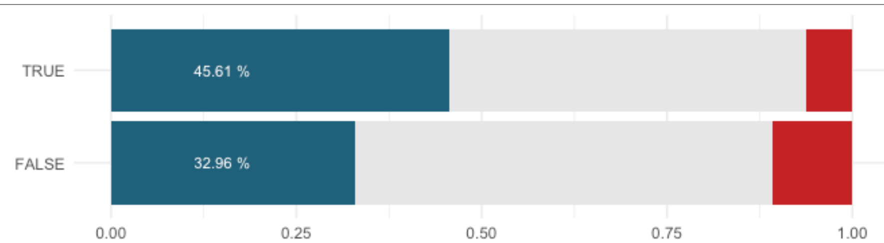
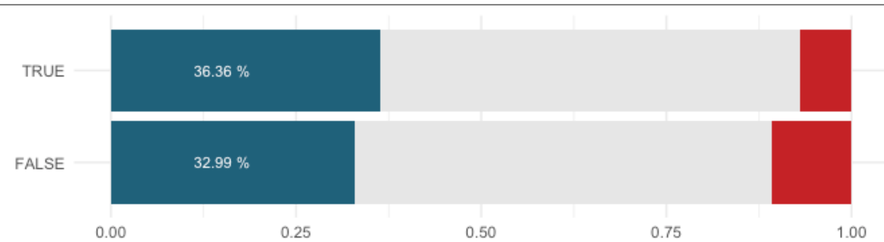
Table 44. Sentiments by activity on the 606

Activity	Sentiments for each activity (True: Doing activity, False: Not doing)
Physical	

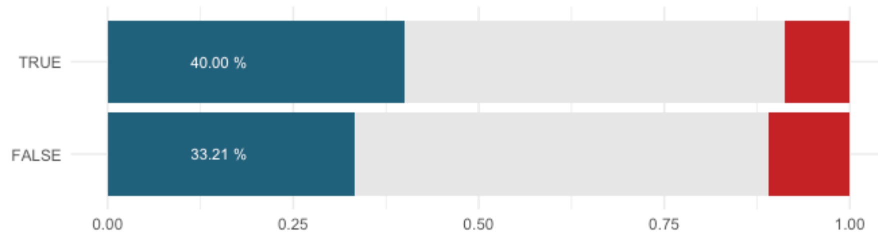


High Line users post more positively about picnics and overlooking experiences than 606 users (45.61% vs. 44.44%). Users of the 606 tweet positively about picnics slightly more often than about unrelated subjects (44.44% vs. 42.40%). However, the High Line shows a greater difference here (45.61% vs. 32.96%). This suggests that it offers more places to have lunch or snacks than the 606. One surprising result is about viewing from overlooks. High Line users tweeted positively about viewing more than 606 users (36.36% vs. 25.00%). The 606 users expressed satisfaction with viewing less than about other topics related to the park (25.00% vs. 42.21%). and the overlooking of the High Line (36.36%). Negative tweets about viewing on the 606 are also 25.00%, meaning 606 users are not attracted to the park's views. The overall results suggest that High Line users like picnicking and enjoying the view more than 606 users.

Table 15. Sentiments by activity on the High Line.

Activity	**Sentiments for each activity (True: Doing activity, False: Not doing)**
Physical	
Arts	
Social	
Picnic	
Overlooking	

Nature



2.4 Summary

These results suggest that (1) the 606 is closely related to residents' daily lives, (2) the important factors in the High Line's success are art and New York City, and (3) the High Bridge Trail offers diverse opportunities for recreation. Users mainly post about the 606 in May, July, September, and October, and the largest number mention physical activities such as cycling and walking, which were among the goals in the park's design. Interestingly, users frequently discuss affordable housing issues and crime on the 606, sharing crime news and protesting housing prices and taxes on social media, suggesting that these are issues users are interested regarding the park near their homes.

There are two major factors in the High Line's success. Users tweet the most about the park from March to October, suggesting the park's usage is not much related to the season. One interest result is that "art" and "New York City" are frequently mentioned along with "High" and "Line," more than with the other two parks. Previous research has focused on how the park affects economic growth (Ascher & Uffer, 2015) and not on the background to its success. But these two keywords can be considered to reveal one reason for its popularity. Another finding involves the terms "scenic" and "bridge." The bigram analysis links these with "High" and "Line," though they can also be considered part of the previous finding, on "New York City," as the views from the High Line show New York City. The park's observation points may be important because they let users view the city. These findings support the claim that the High

Line's success results from its location and its arts-focused strategy.

The High Bridge provides various recreational opportunities such as biking, walking, and riding. Although the collected data on the High Bridge are not sufficient to produce results, cluster and sentiment analysis show the following: (1) The High Bridge is used for various activities including walking, riding, cycling, photography, and travel. (2) High Bridge tweets describe the background of the trail with "unused," "railroad," "donated," "thirty," "one," and "miles." (3) High Bridge Trail users' postings have become more positive since 2016.

3. How Does Social Media Data Differ from Survey Data?

3.1 Who uses social media, and why?

We now compare several survey respondents and social media users to gauge differences in their populations and the limitations of the two methods. Questions about social media usage were included in the survey, such as "Will you post about today's experience in the park on social media?" and "Why do you post on social media?" These questions provided information about numbers of postings and the reasons for them. Figure 28 shows these: 43% of 606 users, 39% of High Line users, and 28% of High Bridge users expressed the intention to post their experiences. The overall mean was 39%, however, which is much higher than the actual percentage of posts, given the parks' annual visitor numbers. The average percentages of actual postings, based on total visitors, are shown in Table 16. These results may indicate that people who participated in the survey were more likely to share their activities on social media.

Reasons for posting are also shown in Figure 28. High Line respondents tended to post their sentiments more often than the other two parks' respondents. The High Line users said that they would describe their experiences positively (43%) more often than negatively (6%). The

606 respondents said the same, in lower percentages (25% to 2%). These results are similar to the results of the sentiment analysis. As previous studies have found (Brkovic & Stetovic, 2013), social media users and park visitors tend to post their positive emotions and satisfying experiences on social media.

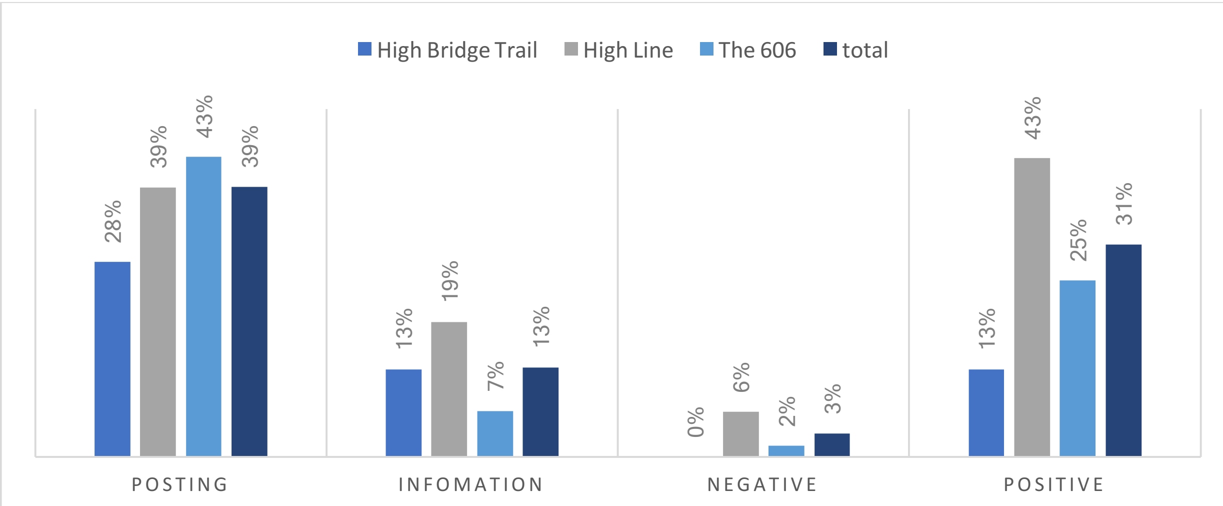


Figure 28. Social media usage of survey respondents.

Table 16. Comparison of posting percentages.

Linear parks	Numbers of tweets/total visitors	Percentage of postings[1]
The 606	4.09% (14,340/3,500,000)	43%
High Line	10.31% (206,229/20,000,000)	39%
High Bridge	0.005% (457/900,000)	28%

[1] Respondents who answered that they would post about their experiences in the parks.

3.2 How are park visitor activities portrayed differently in social media data than in survey data?

To compare surveys and social media analytics, we compare their portrayals of activities in and benefits of linear parks. Five activities are identified from survey and social media data: physical activities, social interaction, overlooking, art viewing, and picnicking. For each, the rate of participation according to the survey and the number of mentions in tweets are computed so that their orders can be compared between the parks. Three benefits are used for the comparison: interactions with nature, social interactions, and improved health. To identify the perceived benefits of each activity, we filter the data using keywords related to each activity and then analyze sentiments; positive sentiments indicate that users were satisfied with their experiences. After identifying the benefits, we compare their order between the parks.

Table 17. Comparison of order of activities between parks

	Physical activity		**Social interaction**		**Art viewing**		**Overlooking**		**Picnic**	
	s[1]	**sm**[2]	**s**	**sm**	**s**	**sm**	**s**	**sm**	**s**	**sm**
The 606	1	3	**2**	**2**	**2**	**2**	3	2	**2**	**2**
High Line	3	2	**1**	**1**	**1**	**1**	**1**	**1**	1	3
High Bridge	2	1	**3**	**3**	**3**	**3**	2	3	3	1

[1] Survey result (number of respondents who did the activity/total number of respondents).

[2] Social media data result (number of tweets mentioning keywords/total number of tweets).

Regarding activities, the social media and survey data show similar results for social interaction and art viewing, and somewhat similar results for physical activity and overlooking. There is no similarity in picnicking. Table 17 compares the results of the two methods. Social interaction and art viewing have the same order with both methods. High Line users are the most active with friends and family, followed by 606 users; High Bridge users rarely mention friends or family or visit with them by comparison. The same order appears for art viewing: High Line, 606, and then High Bridge.

High Line respondents enjoy scenic overlooks more than the other parks' users, but the survey and social media data differ after that: social media analytics indicate that 606 users enjoy overlooks more than High Bridge respondents, but the surveys report the opposite. The same result appears for physical activities: the survey data say that 606 users participate in physical activity the most, but the social media data say they do so the least.

In picnicking, 606 users fall in the middle with both methods. These results suggest that social media and survey data partly produce the same results .

3.3 How does park visitors' satisfaction appear differently in social media and in surveys?

I compare the order of perceived benefits from the survey and from percentages of positive tweets (Table 18). Social and natural benefits have the same orders in both sources. Because the High Bridge social media results are 0 for natural benefits and 100% for social benefits, these are removed to reduce error. 606 users report more social and natural benefits than High Line users, but health benefits have a different result: Survey respondents gain the most health benefits from the 606 and then the High Bridge, but in social media analytics the High Bridge is first and the 606 second. (High Line users report the fewest either way). These results show that social media analytics can have the same results as surveys in terms of benefits.

Overall, these results indicate that social media analytics and surveys have similar implications regarding benefits and activities. Activities appear in the same order, except for physical activities and picnics. Social media analytics can thus substitute for surveys with other activities. The same is true with benefits: two of the three kinds of benefits have the same order with both methods. These results support the use of social media data in place of surveys to compare parks, though they cannot replace surveys altogether.

Table18. Comparison of the order of perceived benefits.

	Health benefit		**Social benefit**		**Natural benefit**	
	s[1]	**sm**[2]	**s**	**sm**	**s**	**sm**
The 606	1	2	**1**	**1**	**1**	**1**
High Line	**3**	**3**	**2**	**2**	**2**	**2**
High Bridge	2	1	-	-	-	-

[1] Survey result (mean of benefit factors).

[2] Social media result (positive tweets mentioning keywords/total tweets mentioning keywords).

4. How Do Landscape Architects use Social Media and Surveys?

4.1 In what situation might each type of data be best used?

The results so far are summarized in Table 19. The three parks studied here have different environments. For example, the 606 is in a residential area, and the High Line crosses the center of a large city. The parks also differ in number of visitors. The difficulty of the data collection is summarized on the basis of the process and the results for both surveys and social media.

The table also recommends appropriate kinds of data for each context. For the 606, both surveys and social media are recommended. The park does not have a high proportion of tourists and is used mainly by residents. Enough data were collected for social media analysis, and this is particularly suitable for a space for daily activities, like the 606, which requires a time series analysis from the past to the present. For the High Line, social media analysis is more effective and efficient than surveys. Surveys are difficult because the many foreign tourists create language barriers; however, adequate social media data are available for the park for people of multiple nationalities.

Table 19. Recommendations for selecting data sets

Linear parks	The 606	The High Line	The High Bridge
Context			
Neighborhoods	Residential area	Urban and commercial area	Rural area
Design features	Physical activity	Public arts and being in nature	Being in nature and physical activity
Number of visitors	1,000,000/year	7,000,000/year	200,000/year
Survey			
Number of responses	320	310	120
Response rate	63.12%	42.18%	84.51%
Survey period	7 days	5 days	4 days
Difficulty of survey	Moderate	Difficult	Moderate
Social Media Data			
Number of tweets (01/2015–06/2019)	14,340	206,229	457

Difficulty	Easy	Easy	Easy
Recommendation			
Survey	Recommend	Recommend	Highly Recommend
Social media data	Recommend	Highly Recommend	Recommend

The High Bridge does not have enough social media data to make social media analysis possible. Surveys are better in cases like this and where there are few users.

Social media analytics probably cannot replace surveys completely, as it is difficult to secure users' information, but it is a valuable approach for being able to collect unanticipated information. Table 20 compares surveys and social media analysis. Surveys are useful for examining users' basic characteristics, their behavior in parks, and the benefits and satisfaction they receive. Social media analysis is useful for capturing park activities, people's interests, and benefits. In particular, it is also valuable for being able to use past information. As a result, social media analysis is an appropriate tool for identifying the overall usage of parks and interests of their users before surveys are conducting surveys, whereas more detailed information must be gathered through surveys.

Table 20. Comparison of two data sets for understanding park users

User characteristics	Survey	Social media
User		
Demographics	●	○
Activities	●	●
Interests	○	●
Benefits and **satisfaction**	●	●
Time		
Past	○	●
Present	●	●

CHAPTER V. CONCLUSION

The purpose of this study was (1) to identify visitors' behaviors in and perceptions of linear parks, (2) to identify social media users' behaviors in and perceptions of linear parks, and (3) to compare small data with big data. This chapter discusses the main findings and their implications for practitioners such as landscape architects and urban planners. It has three sections. The first addresses the main findings in the order of the research questions at the center of the study. The second describes implications and recommendations for practitioners. The final section discusses the limitations of the study and suggests directions for future work.

This study used two main analytics: statistical analytics for the survey, and text mining for the social media data. The data were obtained from park visitor surveys and social media posts that mentioned linear parks. Four questions and nine sub-questions guided the study:

Research Question 1: What information can landscape architects obtain from surveys?

1-1: Who responds to surveys?

1-2: What park visitor activities can be investigated using surveys?

1-3: What park visitor benefits can be investigated using surveys?

Research Question 2: What information can landscape architects obtain from social media analytics?

2-1: What park visitor activities can be investigated using social media analytics?

2-2: What park visitor interests can be investigated using social media analytics?

2-3: What park visitor benefits can be investigated using social media analytics?

Research Question 3: How do social media data differ from survey data?

3-1: How are park visitor activities portrayed differently in social media and survey data?

3-2: How is park visitor satisfaction portrayed differently in social media and survey data?

3-3: How does park visitors' satisfaction appear differently in social media and in surveys?

Research Question 4: How can landscape architects use social media analytics and surveys?

4-1: In what situation might each type of data be best used?

1. Summary of Major Findings

This section summarizes the answers to the research questions, in three parts. The first part is about the findings from social media analytics. Based on three components, activities, opinions, and satisfaction, three sub-findings are reported. The second section gives the results of the comparison between surveys and social media analytics. The third section presents implications for practitioners by answering research question 3.

1.1 Findings from research question 1

For phase one, two surveys were conducted: the park visitors' survey and a park survey. The visitors' survey was designed to understand who visits linear parks, what they do in them, and what they think of the parks. Respondents' characteristics were described before the statistical analytics, to identify who uses linear parks. The survey revealed two components of the landscape perception model: people's activities and their attitudes. For the last component of the model, physical features, the park survey was used to identify the uniqueness and connectivity of each park as physical features.

1.1.1 Finding 1: Visitor characteristics in linear parks

Before the findings could be reviewed on the basis of the research model, it was necessary to compare the respondents in the three parks to understand their visitors. Visitors to the 606 visitors are young (age 18–39 = 59.37%), highly educated (graduate school = 32.81%), and high-income (>$75,000 = 54.38%). They visit the park frequently (2–5 days a week = 35%) by walking (64.06%) from nearby (1–2 miles = 39.69%) and spend thirty minutes to an hour there (49.69%). High Line visitors are also young (age 18–39 = 55.17%), educated (university = 57.54%), and middle-income (25,000–100,000 = 46.13%). They visit less than once a month (35.48%) by taking public transportation (59.03%) from 3 to 20 miles away (38.06%) and spend thirty minutes to an hour in the park (54.19%). High Bridge visitors are older (age 30–59 = 52.50%), mostly white (80%), high-income (>$75,000 = 49.16%), and university educated (43.33%). Most are first-time visitors (42.50%) arriving by car (62.50%) from 11 miles or more away (52.50%) and spend over an hour (56.67%) in the park.

These results indicate that the 606 is highly used by locals who visit daily; that the High Line is used more occasionally than daily, and that the High Bridge Trail is used mainly by older and retired tourists rather than local residents.

1.1.2 Finding 2: Visitor activities in linear parks, according to the survey

The major activity in all three linear parks is physical activity. This agrees with previous research (Henderson, 2006; Kaczynski et al., 2014; Keith et al., 2018). Many visitors to the 606 engage in biking (48.13%), walking (80.94%), running (28.75%), and dog walking (17.81%). The High Bridge has the next-most visitors, and they also engage in physical and nature activities (49.17%). Visitors to the High Line also engage in physical activities, but mainly just walking (76.66%; running = 10.97%). Overall, then, people visit the 606 to improve their health, mainly through walking and biking, people visit the High Bridge visitors to experience nature, and they also engage in physical activities, and the High Line visitors aim to explore nature but enjoy public art more (40.97% vs. 13.87%).

Table 21 summarizes the order of visitors' activities in each park. Physical activities are the most popular, followed by viewing at observation points and relaxing. However, people rarely participate in social activities such as picnicking. Visitors to the High Line take part in diverse activities, including social interaction, relaxation, and viewing, which could mean that the High Line provides more varied experiential places than the other two parks. The 606 is used for physical activity more than the other two. Because it was designed for this, the present study shows that in- design strategies work. Visitors to the High Bridge Trail have the least varied activities; They don't engage in social interaction or relaxation.

Table 21. Activities in linear parks

Activities	The 606	High Line	High Bridge
Physical	1	3	2
Social	2	1	3
Relaxation	2	1	3
Viewing (overlooks)	3	1	2

1.1.3 Finding 3: Visitor benefits from linear parks, according to the survey

Finding 2 involves visitors' attitudes toward linear parks, which can be divided into two components: perceived benefits and satisfaction. Three benefits were identified from the survey responses: health benefits, natural benefits, and social benefits. Table 22 shows the order of these. 606 visitors report receiving health benefits the most (mean = 4.51), and then natural (3.89) and social benefits (3.53). High Line visitors report these in the same order. High Bridge visitors report natural benefits (4.17) more than the other two parks and health benefits slightly less (4.47) than 606 users (4.51).

Table 22. Perceived Benefits

Perceived benefits	The 606	High Line	High Bridge
Health benefits	1	3	2
Nature benefits	2	3	1

Social benefits	1	3	2

Overall, visitors to all three parks perceive the benefits in the same order. This result supports previous studies that verified physical activities and health benefits in linear parks (Henderson, 2006). Among the three parks, visitors to the 606 perceive the most benefits, and visitors to the High Line perceive the fewest —in spite of the park's popularity, interestingly.

1.2 Finding from social media analytics

This section summarizes the results of the social media analytics: clustering, sentiment analysis, and bigram network analysis were used to answer the research questions.

1.2.1 Finding 4: Social media users' activities

To identify social media users' behaviors in linear parks, we used keyword searching and calculated the percentages of tweets to mention each. Then we conducted a topic analysis by month. The keyword search identified the main kinds of activities, and these were compared by year and month. Because the tweets about the High Bridge were too few for keyword searching and topic analysis, the High Bridge was excluded from these analytics.

By year, social and activities show similar patterns from 2015 to the present. Since 2017, both art and physical activities have increased. An interesting pattern appears in the monthly analysis: in July, the High Line has the lowest rate of physical activity but the highest rates of art and social activities. This may be because in the hot summer, visitors tend to view public arts instead of engaging in physical activity.

The topic analysis for the High Line shows that "art," "photo," "city," and "walk" are mentioned in all months, and “people,” “friend,” “love,” and “NY” are mentioned in most. This indicates that the High Line is a landmark rather than just a park. The same topics, such as “art” and “photo,” are also repeated by month and by season for the High Line, which could mean that time of year is not an important factor for that park.

The 606 has similar results. For example, art and social activities show the same pattern as with the High Line, increasing in May and July as physical activity decreases. One difference from the High Line is in physical activity. On the 606, physical activity increases in the winter, from November to January. This could mean that social media users tend to post their physical activities during the harsh winter weather of Chicago.

The topic analysis for the 606 compares interestingly to that of the High Line. In winter, social media users mention “light,” “walk,” “jazz,” and “neighborhood.” In spring, they mention “bike,” “gentrification,” “housing,” and “price.” In summer, they mention “bike,” “run,” and “happiness,” and in fall they mention “photo,” “beautiful,” and “Halloween.” These results, particularly “gentrification,” “robberies,” “housing,” and “price,” show that the 606 is closely related to residents’ personal interests. One interesting topic is physical activity; 606 users mention physical activities differently by season—for example, “light” and “walk” in winter and “bike” in spring. This could mean that 606 users engage in different physical activities each season. And whereas High Line users mention “photo” almost every month, 606 users mention “photo” alongside “beautiful” in the fall. The biggest contrast is in “art”: 606 users do not mention it at all in the topic analysis results.

1.2.2 Finding 5: Social media users’ interests

Text mining revealed several facts about the High Line. The first is that its location, crossing the center of Manhattan, contributes to its success: according to the cluster analysis, which identifies related words, the High Line is always mentioned together with “New York” and “NYC.” Second, “photo,” is often mentioned with “High Line.” It can be inferred that visitors tend to upload photos of the High Line to represent where they are at the moment. Third, the total number of tweets decreases in winter but the sentiments are similar in all season. This also supports the hypothesis that the High Line is a landmark rather than a park.

The results for the 606 show several differences. First, the fewest tweets were uploaded from June to August and the most in September and October. Second, sentiment analysis supports the hypothesis that in June, people post their activities less because of the crowded trail. Third, among positive-sentiment tweets, park users mention “jazz” and “free event” a lot; among negative-sentiment tweets, people frequently mention “robberies” and “gentrification.”

The High Bridge Trail has fewer data, so it is harder to analyze and the findings may be limited. First, there are almost no tweets in November and December, which could mean that users rarely visit in winter. Second, the dendrogram analysis indicates that the park is designed and used for biking, exploring, and hiking. Third, positive sentiment about the High Bridge Trail has increased since 2015, which could mean that users are more satisfied with the park.

1.2.3 Finding 6: Relationship between social media users’ behaviors and attitudes

Table 23. Positive sentiment by activity

Activity	The 606	High Line	High Bridge
Total positive sentiment	42.21%	32.99	45.95
Physical activity	51.78	43.88	64.28
Arts activity	48.02	39.54	66.66
Social activity	89.00	84.59	100.00
Viewing from overlooks	25.00	36.36	50.00

Table 23 describes the results of the sentiment analysis based on activities detected from tweets. Comparing the results of the survey and the social media analytics, we find that they partially support each other.

1.2.4 Summary

First, the 606, which is designed for residential and mixed use, is closely related to residents' daily lives. Users like to bike and walk in the park and are interested in events and worried about housing prices and crime. The clustering analysis shows that biking and walking are the main activities mentioned on social media. As the 606 was designed for these things, (Sinha, 2014), this means it provides well for residents' needs. Interestingly, the cluster analysis also indicates that 606 users are worried crime and gentrification. Users also rarely mention the park in the middle of summer or winter. Because winter in Chicago is too cold for many outside activities, a low volume of tweets is expected then, but the summer case is more interesting.

Second, this study identifies a reason for the High Line's success. Users mention the park in all seasons but November and December, and the number of tweets is ten times that than of the 606. The cluster analysis shows that users frequently mention the New York City and the arts at the same time, and the sentiment analysis and bigram network show that the city and the urban scenery are the topics most often mentioned along with this park, much more so than with the other two. This suggest that the power of New York City lies behind the success of the park, along with the popularity of the arts, which are often mentioned arts with positive sentiment. These results were not detected through the survey.

Third, the High Bridge results show that users engage in various physical activities in the park and are interested in its background. Users mainly mention the park in the summer and rarely in the winter. The main topic of the tweets is the park's origin in unused railways donated to Virginia. Users also frequently mention various activities, such as biking, walking, exploring, traveling, and taking photos. Due to the lack of tweet volume in comparison to the other two parks, the social media data tend to provide more specific information.

Overall, these results let us identify the strengths and weaknesses of social media analytics. Social media analytics can tell us the main issues related to a linear park and help us find hidden topics. For example, cluster analysis contributed to our identification of housing and crime as major issues on the 606, and the city as one of the reasons for the High Line's success. Furthermore, social media data let researchers trace the past by following trends of social media use over time. Finally, these data help us understand the topics users feel positively or negatively about, such as housing prices and crime in the case of the 606.

On the other hand, the High Line users like to look at urban scenery and arts. This supports the use of social media to understand users of linear parks. However, social media analytics are not always suitable for every park. For example, the High Bridge tweets are too specific due to the small volume collected, (only 475 from 2015 to 2019), and the results of the social media analytics thus tend to be exaggerated. If a term is posted only five times, the analytics detects it as among the most frequently used words. Still, however, social media analytics can capture information about parks that surveys miss when the data are sufficient.

1.3 Finding from the comparison

1.3.1 Finding 7: Comparing the two methods

This part concludes the comparison of surveys and social media analytics. Table 24 compares the basic information about the two data sources. For this study, 750 survey responses and 178,756 tweets were collected. The survey was conducted from April to May 2019 in each linear park. Because this had to be done on site and in person, it was costly and laborious. The social media data were collected from the entire period of January 2015 to June 2019 using keyword searches of Twitter. A basic knowledge of programming languages was needed, but otherwise the process was easier, cheaper, and faster than the survey.

Regarding the collected information, the survey requested basic information such as the respondent's age, gender, race, income, and education level; that information is hard to identify in social media data (Ruths & Pfeffer, 2014), so the study does not include the demographics of the social media users. Survey respondents also reported their activities through a designed questionnaire that asked directly about perceived benefits and satisfaction. By contrast, social media data are generated by users; there is no space for a researcher in the data generation.

Table 24. Comparison of the two data sources used in this study

	Survey	**Social media**
Sample size	750 responses	178,756 tweets
Study period	April–May, 2019	January 2015–June 2019
Topics	Demographics: income, race, education	Unknown
	Perceived benefits, satisfaction	Sentiment polarity, sentiment score
	Self-reported activities	Identified activity based on term-frequency
Cost of data collection	Collection at study site: Laborious and costly	Collection through social media platform: cheap

1.3.2 Summary

This study compares two methods of data collection, focused on activities and benefits. The survey asked respondents to check all the activities they did in the park. Social media users' activities were detected by term frequency in social media data. Both results ordered the activities similarly. For example social interaction and art viewing were most popular on the High Line, then the 606, then the High Bridge according to both methods. Both methods also reported that High Line visitors engaged in viewing from overlooks the most. As for benefits, according to both methods vistors to the 606 were more satisfied than High Line visitors with the parks' social and natural benefits. These results suggest social media analytics can replace surveys when the textual information is sufficient for analysis.

Social media analytics also differ from surveys in accuracy of results. For example, social media revealed that 606 users were interested in events and worried about housing prices and crimes, but the pre-designed survey could not capture those facts. Social media analytics can also catch hidden and more general information: through cluster analysis, we found possible reasons for the High Line's success in the arts and in the New York City itself. These results involve general information that would be hard to identify through a survey.

On the other hand, surveys provide specific information and can describe visitors' demographics, motivations, travel information, and specific benefits. For example, 606 users tend to be young, high-income, well educated, white, and female. These data cannot be collected through social media.

2. Recommendations

This study provides recommendations to designers and planners of linear parks. Observation points may be an attractive facility for such parks, but visitors rarely enjoy making use of them, and the activity does not add to their perceived benefits or satisfaction. Nature overlooks only have a positive relationship with natural benefits ($z = 2.11$, $p < .01$). Most visitors to the High Line use the overlooks (70.83%), and they enjoy the scenery of New York City. But visitors prefer physical and social activities such as walking, jogging, chatting, and meeting family or friends.

Linear parks are used mainly for physical activities, as previous research has shown

(Crewe, 2001). Particularly because they are long paths that do not allow cars, many people use these parks to run and bike. Designers of linear parks must decide whether to allow cycling and then carefully make paths accordingly. An important element of linear parks is the circulation plan for pedestrians, runners, and cyclists.

In urban areas, linear parks must also provide places for relaxation and picnics rather than physical activity. Visitors to the High Line report that they liked picnicking, interacting with others, and viewing public art more than visitors to the other parks, through both media. They also take part in physical activity less than the other two parks' users. This suggests that physical activity is not the main function of the urban linear park. Furthermore, the important things about the urban linear park are the city itself and the arts. The social media analytics for the High Line show that users frequently mention New York City and arts together with the park, and the bigram network shows a relationship between scenery and the city at the bridge (an observation point). It can be concluded that one reason for the High Line's success is its location. Planners should recognize that not all linear parks can achieve the same effect.

Linear parks in suburban and residential areas are used mainly for physical activity. For example, 606 visitors mostly engage in walking, jogging, biking, and skateboarding. High Bridge visitors also like to do physical activities there. The long and linear shape of these parks allows visitors to do these activities, which suggests that linear parks in residential areas should be considered spaces for physical activity.

Planners of linear parks in residential areas must also carefully consider policies related to gentrification. People who use the 606 and live nearby worry about housing prices and sometimes protest them. To alleviate these issues, planners must prepare for them.

Finally, connectivity has been considered an important function of bridges, but linear parks in suburban areas are not used to connect points of interest. Visitors to the 606 rarely use it to reach a destination. Rather than increasing connectivity, they want to minimize access points to manage crime in the park.

3. Limitations of the study and implications for future research

The aim of this study was to discover hidden facts about linear parks using small data and big data. One goal was to provide recommendations for designers and planners. Another was to compare big and small data sources to assess the value of big data to landscape architecture.

Surveys, a small data source, have limitations in sample selection. Many people on the High Line spoke no English; because the questionnaire and the researcher used English, these people's responses could not be included. Future researchers must consider tourists' responses.

Social media data and analytics are limited by population bias. Social media users tend to young, white, and high-income (Chou et al., 2009; Ruths & Pfeffer, 2014), and analysis results will represent their thoughts and behaviors. Park visitors, on the other hand, tend to be young, white, and male (Furuseth & Altman, 1991; Kemperman & Timmermans, 2008). It cannot be determined whether further studies can reduce the population bias in social media; unless social media users state their demographic data publicly, this bias cannot be eliminated.

Reference:

11 of the best social media analytics tools for brands. (2019, May 20). Retrieved July 30, 2019, from Sprout Social website: https://sproutsocial.com/insights/social-media-analytics-tools/

Afzalan, N., & Muller, B. (2014). The Role of Social Media in Green Infrastructure Planning: A Case Study of Neighborhood Participation in Park Siting. *Journal of Urban Technology*, *21*(3), 67–83. https://doi.org/10.1080/10630732.2014.940701

Agrawal, D., Budak, C., El Abbadi, A., Georgiou, T., & Yan, X. (2014). Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks. In A. Madaan, S. Kikuchi, & S. Bhalla (Eds.), *Databases in Networked Information Systems* (pp. 1–16). Springer International Publishing.

Ahern, J. (1991). Planning for an extensive open space system: Linking landscape structure and function. *Landscape and Urban Planning*, *21*(1), 131–145. https://doi.org/10.1016/0169-2046(91)90037-M

Ahmad, I. (2018, June 15). How Much Data Is Generated Every Minute? [Infographic]. Retrieved September 15, 2018, from Social Media Today website: https://www.socialmediatoday.com/news/how-much-data-is-generated-every-minute-infographic-1/525692/

Aoki, Y. (1991). Evaluation Methods for Landscapes with Greenery. *Landscape Research*, *16*, 2–6.

Ascher, K., & Uffer, S. (2015). The High Line Effect. *Global Interchanges: Resurgence of the Skyscraper City*, 243–228.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., … Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*(1), 481–518. https://doi.org/10.1140/epjst/e2012-01703-3

Becken, S., Stantic, B., Chen, J., Alaei, A. R., & Connolly, R. M. (2017). Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. *Journal of Environmental Management*, *203*, 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007

Betz, C. J., English, D. B., & Cordell, H. K. (1999). Outdoor recreation resources. *Outdoor Recreation in American Life: A National Assessment of Demand and Supply Trends*, 39–182.

Bighorse, A. (2010, May 13). The Highline: Past and present. Retrieved March 13, 2019, from https://web.archive.org/web/20141023213640/http://www.gsd.harvard.edu/pbcote/courses/archive/2010/gsd6447/bighorse/

Birch, E. L. (1986). *The Observation Man, A Study of William H. Whyte*.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

Blank, G., & Lutz, C. (2016). The Social Structuration of Six Major Social Media Platforms in the United Kingdom: Facebook, LinkedIn, Twitter, Instagram, Google+ and Pinterest. *Proceedings of the 7th 2016 International Conference on Social Media & Society*, 8:1–8:10. https://doi.org/10.1145/2930971.2930979

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brindley, P., Cameron, R. W., Ersoy, E., Jorgensen, A., & Maheswaran, R. (2019). Is more always better? Exploring field survey and social media indicators of quality of urban greenspace, in relation to health. *Urban Forestry & Urban Greening*, *39*, 45–54. https://doi.org/10.1016/j.ufug.2019.01.015

Brkovic, M., & Stetovic, B. (2013). Harnessing social media for urban planning: An overview. *Proceedings of the 49th International Society of City and Regional Planners Congress, Brisbane, Australia*, 1–4.

Broder, A. Z., Glassman, S. C., Manasse, M. S., & Zweig, G. (1997). Syntactic clustering of the Web. *Computer Networks and ISDN Systems*, *29*(8), 1157–1166. https://doi.org/10.1016/S0169-7552(97)00031-7

Carmona, M. (2010). *Public places, urban spaces: The dimensions of urban design*. Routledge.

Carr, S. J., & Lassiter, A. (2017). Big Data, Small Apps: Premises and Products of the Civic Hackathon. In P. (Vonu) Thakuriah, N. Tilahun, & M. Zellner (Eds.), *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 543–559). https://doi.org/10.1007/978-3-319-40902-3_29

Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. *Landscape and Urban Planning*, *170*, 283–292. https://doi.org/10.1016/j.landurbplan.2017.07.004

Cheng, Z., Caverlee, J., & Lee, K. (2010). You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 759–768. https://doi.org/10.1145/1871437.1871535

Chicago Department of, & Transportation. (2012). *Bloomingdale Trail and Park Framework Plan*. City of Chicago.

Chou, W.-Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social media use in the United States: Implications for health communication. *Journal of Medical Internet Research*, *11*(4).

Cohen, B. E., Martires, K. J., & Ho, R. S. (2016). Psoriasis and the Risk of Depression in the US Population: National Health and Nutrition Examination Survey 2009-2012. *JAMA Dermatology*, *152*(1), 73–79. https://doi.org/10.1001/jamadermatol.2015.3605

Compton, R., Lee, C., Lu, T., Silva, L. D., & Macy, M. (2013). Detecting future social unrest in unprocessed Twitter data: "Emerging phenomena and big data." *2013 IEEE International Conference on Intelligence and Security Informatics*, 56–60. https://doi.org/10.1109/ISI.2013.6578786

Cranz, G. (1978). *Changing roles of urban parks: From pleasure garden to open space*. Institute of Urban & Regional Development, University of California, Berkeley.

Crewe, K. (2001). Linear Parks and Urban Neighbourhoods: A Study of the Crime Impact of the Boston South-west Corridor. *Journal of Urban Design*, *6*(3), 245–264. https://doi.org/10.1080/13574800120105779

Crompton, J. L. (2001). The Impact of Parks on Property Values: A Review of the Empirical Evidence. *Journal of Leisure Research*, *33*(1), 1–31. https://doi.org/10.1080/00222216.2001.11949928

David, J. (2002). *Reclaiming the High Line*. Design Trust for Public Space, Incorporated.

David, J., & Hammond, R. (2011). *High Line: The inside story of New York City's park in the sky*. Farrar, Straus and Giroux New York.

De Sousa, C. A. (2006). Unearthing the benefits of brownfield to green space projects: An examination of project use and quality of life impacts. *Local Environment*, *11*(5), 577–600. https://doi.org/10.1080/13549830600853510

Driver, B. L., Brown, P. J., & Peterson, G. L. (Eds.). (1991). *Benefits of leisure*. State College, PA, US: Venture Publishing.

Driver, Beverly L. (1977). *Item pool for scales designed to quantify the psychological outcomes desired and expected from recreation participation*. Rocky Mountain Forest and Range Experiment Station.

Driver, Beverly L, Brown, P. J., & others. (1978). The opportunity spectrum concept and behavioural information in outdoor recreation resource supply inventories: A rationale. In: Integrated inventories of renewable natural resources: Proceedings of the workshop, January 1978, Tucson, Arizona (Edited by HG Lund et al.). *USDA Forest Service, General Technical Report*, (RM-55), 24–31.

Eiter, S., & Vik, M. L. (2015). Public participation in landscape planning: Effective methods for implementing the European Landscape Convention in Norway. *Land Use Policy*, *44*, 44–53. https://doi.org/10.1016/j.landusepol.2014.11.012

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, *102*(3), 571–590. https://doi.org/10.1080/00045608.2011.595657

Evans, L., & Saker, M. (2017). *Location-Based Social Media: Space, Time and Identity*. Springer.

Ewing, R. (2001). Using a visual preference survey in transit design. *Public Works Management & Policy*, *5*(4), 270–280.

Fehrenbacher, J. (2014, September 20). INTERVIEW: Landscape Architect James Corner On NYC's High Line Park. Retrieved March 13, 2019, from https://inhabitat.com/interview-architect-james-corner-on-the-design-of-high-line/

Fellmann, J. D. (1957). Pre-building growth patterns of Chicago. *Annals of the Association of American Geographers*, *47*(1), 59–82.

Ferretti, V., & Degioanni, A. (2017). How to support the design and evaluation of redevelopment projects for disused railways? A methodological proposal and key lessons learned. *Transportation Research Part D: Transport and Environment*, *52*, 29–48. https://doi.org/10.1016/j.trd.2017.02.008

Fisher, D. M., Wood, S. A., White, E. M., Blahna, D. J., Lange, S., Weinberg, A., … Lia, E. (2018). Recreational use in dispersed public lands measured using social media data and on-site counts. *Journal of Environmental Management*, *222*, 465–474. https://doi.org/10.1016/j.jenvman.2018.05.045

Flink, C., Olka, K., Searns, R., & Conservancy, R. R. to T. (2001). *Trails for the Twenty-First Century: Planning, Design, and Management Manual for Multi-Use Trails*. Island Press.

Flippen, B., & McClintock, R. (2014). *High Bridge: "The End of Our Fifteen Months'' Labor."* Friends of High Bridge Trail State Park.

Floyd, M. F., Spengler, J. O., Maddock, J. E., Gobster, P. H., & Suau, L. J. (2008). Park-Based Physical Activity in Diverse Communities of Two U.S. Cities: An Observational Study. *American Journal of Preventive Medicine*, *34*(4), 299–305. https://doi.org/10.1016/j.amepre.2008.01.009

Fredericks, J., & Foth, M. (2013). Augmenting public participation: Enhancing planning outcomes through the use of social media and web 2.0. *Australian Planner*, *50*(3), 244–256. https://doi.org/10.1080/07293682.2012.748083

Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence*, *35*, 237–245.

Friends of High Line. (2019). The Friends of High Line Official Web Page. Retrieved March 13, 2019, from The High Line website: https://www.thehighline.org/history/

Furuseth, O. J., & Altman, R. E. (1991). Who’s on the greenway: Socioeconomic, demographic, and locational characteristics of greenway users. *Environmental Management*, *15*(3), 329–336. https://doi.org/10.1007/BF02393879

Ganesan, K. (2019, February 23). Getting Started with Text Preprocessing for Machine Learning & NLP. Retrieved August 5, 2019, from Kavita Ganesan website: https://kavita-ganesan.com/text-preprocessing-tutorial/

Gehl, J., & Svarre, B. (2013). *How to study public life*. Island Press.

Georgiou, T. (2017). *Understanding the Real World through the Analysis of User Behavior and Topics in Online Social Media* (Ph.D., University of California, Santa Barbara). Retrieved from http://search.proquest.com/pqdtglobal/docview/1892464099/abstract/B7EF368696BA4188PQ/1

Gobster, P. H., Sachdeva, S., & Lindsey, G. (2017). Up on The 606: Understanding the Use of a New Elevated Pedestrian and Bicycle Trail in Chicago, Illinois. *Transportation Research Record*, *2644*(1), 83–91. https://doi.org/10.3141/2644-10

Godbey, G., & Mowen, A. (2011). The Benefits of Physical Activity Provided by Park and Recreation Services: The Scientific Evidence. *Australasian Parks and Leisure*, *14*(1), 26.

Goodspeed, R. (2013). The limited usefulness of social media and digital trace data for urban social research. *Seventh International AAAI Conference on Weblogs and Social Media*, 06–28.

Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., & Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, *38*, 100958. https://doi.org/10.1016/j.ecoser.2019.100958

Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the World Are You? Geolocation and Language Identification in Twitter. *The Professional Geographer*, *66*(4), 568–578. https://doi.org/10.1080/00330124.2014.907699

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). John Wiley & Sons.

Guhathakurta, S., Zhang, G., Chen, G., Burnette, C., & Sepkowitz, I. (2019). Mining Social Media to Measure Neighborhood Quality in the City of Atlanta. *International Journal of E-Planning Research (IJEPR)*, *8*(1), 1–18. https://doi.org/10.4018/IJEPR.2019010101

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*, *72*, 38–50. https://doi.org/10.1016/j.compenvurbsys.2018.01.007

Hawker, M. D. (2010). *Developer’s Guide to Social Programming: Building Social Context*

Using Facebook, Google Friend Connect, and the Twitter API, The. Pearson Education.

Hayward, J. (1989). Urban Parks. In I. Altman & E. H. Zube (Eds.), *Public Places and Spaces* (pp. 193–216). https://doi.org/10.1007/978-1-4684-5601-1_9

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464–472. https://doi.org/10.1016/j.ijinfomgt.2013.01.001

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85. https://doi.org/10.3390/ijgi6030085

Henderson, K. A. (2006). Urban Parks and Trails and Physical Activity. *Annals of Leisure Research*, *9*(3–4), 201–213. https://doi.org/10.1080/11745398.2006.10816431

Hollander, J. B., & Shen, Y. (2017). Using Social Media Data to Infer Urban Attitudes About Bicycling: An Exploratory Case Study of Washington DC. In *Springer Optimization and Its Applications*. *City Networks* (pp. 79–97). https://doi.org/10.1007/978-3-319-65338-9_5

Hu, X., & Liu, H. (2012). Text Analytics in Social Media. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 385–414). https://doi.org/10.1007/978-1-4614-3223-4_12

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Eighth International AAAI Conference on Weblogs and Social Media*. Presented at the Eighth International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109

Jacobs, J. (1961). *The death and life of great American cities*. Vintage Books.

Jendryke, M., Balz, T., & Liao, M. (2017). Big location-based social media messages from China's Sina Weibo network: Collection, storage, visualization, and potential ways of analysis. *Transactions in GIS*, *21*(4), 825–834. https://doi.org/10.1111/tgis.12266

Jiang, B., Li, D., Larsen, L., & Sullivan, W. C. (2016). A Dose-Response Curve Describing the Relationship Between Urban Tree Cover Density and Self-Reported Stress Recovery. *Environment and Behavior*, *48*(4), 607–629. https://doi.org/10.1177/0013916514552321

Kaczynski, A. T., Besenyi, G. M., Stanis, S. A. W., Koohsari, M. J., Oestman, K. B., Bergstrom, R., … Reis, R. S. (2014). Are park proximity and park features related to park use and park-based physical activity among adults? Variations by multiple socio-demographic characteristics. *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1), 146. https://doi.org/10.1186/s12966-014-0146-4

Kang, C. D., & Cervero, R. (2009). From elevated freeway to urban greenway: Land value impacts of the CGC project in Seoul, Korea. *Urban Studies*, *46*(13), 2771–2794.

Kankanhalli, A., Tan, B. C., & Wei, K.-K. (2005). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS Quarterly*, 113–143.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kaplan, R. (1980). Citizen Participation in the Design and Evaluation of a Park. *Environment and Behavior*, *12*(4), 494–507. https://doi.org/10.1177/0013916580124006

Keith, S. J., Larson, L. R., Shafer, C. S., Hallo, J. C., & Fernandez, M. (2018). Greenway use and preferences in diverse urban communities: Implications for trail design and management. *Landscape and Urban Planning*, *172*, 47–59. https://doi.org/10.1016/j.landurbplan.2017.12.007

Kelley, M. J. (2013). The emergent urban imaginaries of geosocial media. *GeoJournal*, *78*(1), 181–203.

Kemperman, A. D. A. M., & Timmermans, H. J. P. (2008). Influence of Socio-Demographics and Residential Environment on Leisure Activity Participation. *Leisure Sciences*, *30*(4), 306–324. https://doi.org/10.1080/01490400802165099

Kim, E., Lee, J.-A., Sung, Y., & Choi, S. M. (2016). Predicting selfie-posting behavior on social networking sites: An extension of theory of planned behavior. *Computers in Human Behavior*, *62*, 116–123. https://doi.org/10.1016/j.chb.2016.03.078

Kim, H. J., Chae, B. K., & Park, S. B. (2018). Exploring public space through social media: An exploratory case study on the High Line New York City. *URBAN DESIGN International*, *23*(2), 69–85. https://doi.org/10.1057/s41289-017-0050-z

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kleinhans, R., Ham, M. V., & Evans-Cowley, J. (2015). Using Social Media and Mobile Technologies to Foster Engagement and Self-Organization in Participatory Urban Planning and Neighbourhood Governance. *Planning Practice & Research*, *30*(3), 237–247. https://doi.org/10.1080/02697459.2015.1051320

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., & Blaschke, T. (2018). Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *ISPRS International Journal of Geo-Information*, *7*(9), 378. https://doi.org/10.3390/ijgi7090378

Kullmann, K. (2011). Thin parks/thick edges: Towards a linear park typology for (post) infrastructural sites. *Journal of Landscape Architecture*, *6*(2), 70–81.

Laniado, D., & Mika, P. (2010). Making Sense of Twitter. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, … B. Glimm (Eds.), *The Semantic Web – ISWC 2010* (pp. 470–485). Springer Berlin Heidelberg.

Lee, R., Wakamiya, S., & Sumiya, K. (2013). Urban area characterization based on crowd behavioral lifelogs over Twitter. *Personal and Ubiquitous Computing*, *17*(4), 605–620.

Lee, S.-H., Graefe, A. R., & Li, C.-L. (2007). The Effects of Specialization and Gender on Motivations and Preferences for Site Attributes in Paddling. *Leisure Sciences*, *29*(4), 355–373. https://doi.org/10.1080/01490400701394873

Levere, M. (2014). The Highline Park and Timing of Capitalization of Public Goods. *Working Paper*.

Li, J. (2019). Future Research Method of Landscape Design Based on Big Data. *International Conference on Applied Human Factors and Ergonomics*, 92–100. Springer.

Lim, K. H., Lee, K. E., Kendal, D., Rashidi, L., Naghizade, E., Feng, Y., & Wang, J. (2019). Chapter 6—Understanding sentiments and activities in green spaces using a social data–driven approach. In A. Visvizi & M. D. Lytras (Eds.), *Smart Cities: Issues and Challenges* (pp. 77–107). https://doi.org/10.1016/B978-0-12-816639-0.00006-5

Lin, Y., & Geertman, S. (2019). Can Social Media Play a Role in Urban Planning? A Literature Review. In S. Geertman, Q. Zhan, A. Allan, & C. Pettit (Eds.), *Computational Urban Planning and Management for Smart Cities* (pp. 69–84). https://doi.org/10.1007/978-3-030-19424-6_5

Lindsey, G., Wilson, J., Yang, J. A., & Alexa, C. (2008). Urban Greenways, Trail Characteristics and Trail Use: Implications for Design. *Journal of Urban Design*, *13*(1), 53–79. https://doi.org/10.1080/13574800701804033

Liu, B. (2015). *Sentiment Analysis, Opinion Mining, Emotion Classification*. Cambridge University Press.

Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2015). Real-time Rumor Debunking on Twitter. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1867–1870. https://doi.org/10.1145/2806416.2806651

Lomborg, S., & Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *The Information Society*, *30*(4), 256–265. https://doi.org/10.1080/01972243.2014.915276

Mannila, L., Peltomäki, M., & Salakoski, T. (2006). What about a simple language? Analyzing the difficulties in learning to program. *Computer Science Education*, *16*(3), 211–227. https://doi.org/10.1080/08993400600912384

Martin, J. H., & Jurafsky, D. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson/Prentice Hall Upper Saddle River.

Marwick, A. E., & boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Maynard, D., & Funk, A. (2012). Automatic Detection of Political Opinions in Tweets. In R. García-Castro, D. Fensel, & G. Antoniou (Eds.), *The Semantic Web: ESWC 2011 Workshops* (pp. 88–99). Springer Berlin Heidelberg.

McKenzie, T. L., Cohen, D. A., Sehgal, A., Williamson, S., & Golinelli, D. (2006). System for Observing Play and Recreation in Communities (SOPARC): Reliability and Feasibility Measures. *Journal of Physical Activity and Health*, *3*(s1), S208–S222. https://doi.org/10.1123/jpah.3.s1.s208

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. *Fifth International AAAI Conference on Weblogs and Social Media*. Presented at the Fifth International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816

More, T. A., Stevens, T., & Allen, P. G. (1988). Valuation of urban parks. *Landscape and Urban Planning*, *15*(1–2), 139–152.

Myslín, M., Zhu, S.-H., Chapman, W., & Conway, M. (2013). Using Twitter to Examine Smoking Behavior and Perceptions of Emerging Tobacco Products. *Journal of Medical Internet Research*, *15*(8), e174. https://doi.org/10.2196/jmir.2534

Norfolk Southern. (2018). *Annual Reports*. Retrieved from http://localhost:4503/content/nscorp/en/investor-relations/financial-reports/annual-reports.html

O'Riordan, S., Feller, J., & Nagle, T. (2016). A categorisation framework for a feature-level analysis of social network sites. *Journal of Decision Systems*, *25*(3), 244–262. https://doi.org/10.1080/12460125.2016.1187548

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*, *94*, 74–86. https://doi.org/10.1016/j.ecolind.2017.02.009

Parrott, W. G. (2001). Emotions in social psychology: Volume overview. *Emotions in Social Psychology*, 1–19.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.

Peters, K., Elands, B., & Buijs, A. (2010). Social interactions in urban parks: Stimulating social cohesion? *Urban Forestry & Urban Greening*, *9*(2), 93–100. https://doi.org/10.1016/j.ufug.2009.11.003

Poorthuis, A., & Zook, M. (2017). Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. *Journal of Urban Technology*, *24*(4), 115–135. https://doi.org/10.1080/10630732.2017.1335153

Rail-Trail History. (2019). Bloomingdale Trail (The 606) Rail-Trail History. Retrieved August 1, 2019, from https://www.traillink.com/trail-history/bloomingdale-trail-(the-606)/

Revolvy. (2019). "High Bridge Trail State Park" on Revolvy.com. Retrieved August 2, 2019, from https://www.revolvy.com/page/High-Bridge-Trail-State-Park?smv=141945

Roberts, H. V. (2017). Using Twitter data in urban green space research. *Applied Geography*, *81*, 13e20.

Rogers, R. (2013). *Digital methods*. MIT press.

Rowe, G., & Frewer, L. J. (2005). A Typology of Public Engagement Mechanisms. *Science, Technology, & Human Values*, *30*(2), 251–290. https://doi.org/10.1177/0162243904271724

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063–1064. https://doi.org/10.1126/science.346.6213.1063

Sather-Wagstaff, J. (2008). Picturing experience: A tourist-centered perspective on commemorative historical sites. *Tourist Studies*, *8*(1), 77–103. https://doi.org/10.1177/1468797608094931

Sawislak, K. (1995). *Smoldering City: Chicagoans and the Great Fire, 1871-1874*. University of Chicago Press.

Schroeter, R., & Houghton, K. (2011). Neo-planning: Location-based social media to engage Australia's new digital locals. *Australian Planner*, *48*(3), 191–202.

Schweitzer, L. (2014). Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter. *Journal of the American Planning Association*, *80*(3), 218–238. https://doi.org/10.1080/01944363.2014.980439

Seoul Metropolitan Government. (2018). Seoul Solution [Text]. Retrieved September 15, 2018, from Seoul Solution website: https://seoulsolution.kr/ko/node

Shaw, E. (2017). *Parsing Perceptions of Place: Locative and Textual Representations of Place Émilie-Gamelin on Twitter* (Masters, Concordia University). Retrieved from https://spectrum.library.concordia.ca/983426/

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, *142*, 198–211.

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*, *55*, 9–21. https://doi.org/10.1016/j.cities.2016.03.013

Silva, M. D. F. M. e, Calijuri, M. L., Sales, F. J. F. de, Souza, M. H. B. de, & Lopes, L. S. (2014). Integration of technologies and alternative sources of water and energy to promote the sustainability of urban landscapes. *Resources, Conservation and Recycling*, *91*, 71–81. https://doi.org/10.1016/j.resconrec.2014.07.016

Simons, R., & Jaouhari, A. (2001). Local government intervention in the brownfields arena. *Economic Development Commentary*, *25*(3), 12–18.

Sinha, A. (2014). Slow landscapes of elevated linear parks: Bloomingdale Trail in Chicago. *Studies in the History of Gardens & Designed Landscapes*, *34*(2), 113–122. https://doi.org/10.1080/14601176.2013.830428

Smith, A., & Brenner, J. (2012). Twitter use 2012. *Pew Internet & American Life Project*, *4*.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, *78*(2), 319–338. https://doi.org/10.1007/s10708-011-9438-2

Stodolska, M., Shinew, K. J., & Li, M. Z. (2010). Recreation Participation Patterns and Physical Activity among Latino Visitors to Three Urban Outdoor Recreation Environments. *Journal of Park & Recreation Administration*, *28*(2).

Strengell, N., & Sigg, S. (2018). Local emotions—Using social media to understand human-environment interaction in cities. *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 615–620. https://doi.org/10.1109/PERCOMW.2018.8480364

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748. https://doi.org/10.1080/13658816.2011.604636

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1555–1565.

Tasse, D., & Hong, J. I. (2017). Using User-Generated Content to Understand Cities. In *Springer Geography*. *Seeing Cities Through Big Data* (pp. 49–64). https://doi.org/10.1007/978-3-319-40902-3_3

the 606. (2019). The 606 History. Retrieved August 1, 2019, from The 606 website: https://www.the606.org/about/history/

the City of Chicago. (2004). Logan Square Open Space Plan. Retrieved August 1, 2019, from https://www.chicago.gov/content/city/en/depts/dcd/supp_info/logan_square_openspacepl an.html

Tieskens, K. F., Van Zanten, B. T., Schulp, C. J. E., & Verburg, P. H. (2018). Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landscape and Urban Planning*, *177*, 128–137. https://doi.org/10.1016/j.landurbplan.2018.05.002

Troped, P. J., Whitcomb, H. A., Hutto, B., Reed, J. A., & Hooker, S. P. (2009). Reliability of a Brief Intercept Survey for Trail Use Behaviors. *Journal of Physical Activity and Health*, *6*(6), 775–780. https://doi.org/10.1123/jpah.6.6.775

Tu, W., Cao, J., Yue, Y., Shaw, S.-L., Zhou, M., Wang, Z., … Li, Q. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, *31*(12), 2331–2358. https://doi.org/10.1080/13658816.2017.1356464

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*(3), 201–230. https://doi.org/10.1016/S0272-4944(05)80184-7

Waldheim, C. (2016). *Landscape as urbanism: A general theory*. Princeton University Press.

Wang, D., Brown, G., Liu, Y., & Mateo-Babiano, I. (2015). A comparison of perceived and geographic access to predict urban park use. *Cities*, *42*, 85–96. https://doi.org/10.1016/j.cities.2014.10.003

Wang, R., Zhao, J., & Liu, Z. (2016). Consensus in visual preferences: The effects of aesthetic quality and landscape types. *Urban Forestry & Urban Greening*, *20*, 210–217. https://doi.org/10.1016/j.ufug.2016.09.005

Wang, Z., Joo, V., Tong, C., & Chan, D. (2014). Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data. *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*, 899–904. https://doi.org/10.1109/CloudCom.2014.40

Whiting, J. W., Larson, L. R., Green, G. T., & Kralowec, C. (2017). Outdoor recreation motivation and site preferences across diverse racial/ethnic groups: A case study of Georgia state parks. *Journal of Outdoor Recreation and Tourism*, *18*, 10–21. https://doi.org/10.1016/j.jort.2017.02.001

Whyte, William H. (2009). *City: Rediscovering the Center*. University of Pennsylvania Press, Incorporated.

Whyte, William Hollingsworth. (1980). *THE SOCIAL LIFE OF SMALL URBAN SPACES.* Retrieved from https://trid.trb.org/view/521122

Williams, D. R., Schreyer, R., & Knopf, R. C. (1990). The Effect of the Experience Use History on the Multidimensional Structure of Motivations to Participate in Leisure Activities.

Journal of Leisure Research, *22*(1), 36–54. https://doi.org/10.1080/00222216.1990.11969813

Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-In Data. *PLOS ONE*, *9*(5), e97010. https://doi.org/10.1371/journal.pone.0097010

Wu, X., Wood, S. A., Fisher, D., & Lindsey, G. (2017). Photos, tweets, and trails: Are social media proxies for urban trail use? *Journal of Transport and Land Use*, *10*(1). https://doi.org/10.5198/jtlu.2017.1130

Yesmail. (2015). *Lessons learned: How retailers utilized social media in 2014, trends and takeaways for 2015*.

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Morgan & Claypool.

Zhan, X., Ukkusuri, S. V., & Zhu, F. (2014). Inferring Urban Land Use Using Large-Scale Social Media Check-in Data. *Networks and Spatial Economics*, *14*(3), 647–667. https://doi.org/10.1007/s11067-014-9264-4

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, *1*(1), 43–52. https://doi.org/10.1007/s13042-010-0001-0

Zikopoulos, P., Deroos, D., Parasuraman, K., Deutsch, T., Corrigan, D., & Giles, J. (2013). *Harness the power of big data: The IBM big data platform*. McGraw-Hill New York, NY.

Appendix A: IRB Approval Letter

VIRGINIA TECH.

Division of Scholarly Integrity and
Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE:	April 17, 2019
TO:	Patrick Miller, Cemetrius Lynell Bohannon, Thomas Stefan Skuzinski
FROM:	Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)
PROTOCOL TITLE:	Implications for Elevated Park: Learning from Survey and Big Data Analytics
IRB NUMBER:	**19-214**

Effective April 16, 2019, the Virginia Tech Human Research Protection Program (HRPP) and Institutional Review Board (IRB) determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104(d) category(ies) 2(i),2(ii),4.

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a new request to the IRB for a determination.

This exempt determination does not apply to any collaborating institution(s). The Virginia Tech HRPP and IRB cannot provide an exemption that overrides the jurisdiction of a local IRB or other institutional mechanism for determining exemptions.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Determined As:	**Exempt, under 45 CFR 46.104(d) category(ies) 2(i),2(ii),4**
Protocol Determination Date:	**April 16, 2019**

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

Appendix B: Survey Protocol

1. Every fifth person over age 18 will be approached by the survey administrator and asked to take a voluntary survey. During this introduction, the administrator will outline the purpose of the study and the procedures to be followed:

> This is a Virginia Tech research study and it aims to investigate park user characteristics and explore the relationship between types and user characteristics. This survey will take 5 to 10 minutes. This survey has no more than minimal risk to participating, and you may stop the survey at any point without penalty or loss of benefits. By participating in this survey, you may have a chance to understand what landscape architects do to manage a park and to contribute to providing information to the public to guide future policy. If you're willing to participate in this survey, please read the front page, about the consent information. By completing and returning the survey, you indicate your consent to participate in this research.

2. *If the subject declines*, this is the end of the interaction. The survey administrator will note (1) the reason the subject did not respond and (2) the subject's gender, race, ethnicity, and approximate age on the survey cover sheet before approaching the next person.

3. *If the subject accepts*, the subject will be given a clipboard with a pencil and a survey attached. Before the subject starts the survey, the administrator will explain the consent form. Then the administrator will give the subject some time to complete the survey. The administrator will remain in the general area, approaching other people and answering questions as necessary. The administrator will return to collect the survey and answer any final questions after 5 to 10 minutes.

4. After the survey is completed, the subject will be thanked for his or her participation. If the subject requests contact information, the survey administrator will give a business card to the subject. There will not be any follow-up.

Appendix C: Survey Questionnaire

VIRGINIA TECH

Parks Visitor Survey

This study is being conducted in order to better understand the visitors of the elevated park for better planning and design. Your help is voluntary and responses are anonymous and confidential.

1. What do you enjoy most about the park? (specify)____________________

2. What do you least like about the park? (specify)____________________

3. How often do you come to the park?
   ❑ Everyday ❑ Once a week ❑ less than once a month
   ❑ 2-5 days a week ❑ 1-3 times a month ❑ This is the first time

4. Approximately how much total time did you or will you spend in the park today?
   ❑ Just passing through (1-15min.) ❑ 15 - 30 minutes ❑ 30 - 60 minutes ❑ 1-3 hours ❑ Over 3hours

5. Who did you come to the park with today? (check ALL that apply)
   ❑ Alone ❑ Friends ❑ Family ❑ Organized group ❑ Dog(s) ❑ Other (specify)__________

5a. How many people are in this group including yourself? ____________________

6. Approximately how many miles did you travel to visit the park today from your departure point?
   ❑ 1/4 Mile ❑ 1/2 Mile ❑ 1 - 2 Miles ❑ 3 - 10 Miles ❑ 11 - 20 Miles ❑ Over 21 Miles

6a. How do you get to the park? (check **ALL** that apply)
   ❑ Walk ❑ Bike ❑ Car ❑ Public transportation ❑ Other (specify)__________

6b. How easy is it for you to get to the park from your departure point?
   ❑ Difficult ❑ Somewhat Difficult ❑ Neutral ❑ Fairly Easy ❑ Easy

7. Do you use the park as a route to access other places adjacent to the park? ❑ Yes ❑ No

7a. Please check **ALL** places you have visited or will be visiting today. (check **ALL** that apply)
   ❑ Work ❑ Shopping (restaurant…) ❑ Tourist spot (historical sites…)
   ❑ School ❑ Other parks ❑ Other (specify)__________

8. Please check **ALL** the activities you participated in during your visit to the park today. (check **ALL** that apply)
   ❑ Biking ❑ Picnicking ❑ Relaxing/no main activity
   ❑ Hiking/Walking ❑ Viewing from overlooks ❑ Taking photos
   ❑ Jogging/Running ❑ Viewing public arts ❑ Posting on social media
   ❑ Walking a dog ❑ Viewing wildlife ❑ Other (specify)__________

9. Do you share things on social media about the places you visited? ❑ Yes ❑ No

9a. If you post about your visit to the park today on social media, what things will you post? (check **ALL** that apply)
   ❑ Practical information useful to others ❑ Scenery that you like
   ❑ Positive experiences with other people ❑ Things that are unpleased or you dislike
   ❑ Negative experiences with other people ❑ Other(specify)__________

9b. If you post today's experience on social media, whom do you want to share? (check **ALL** that apply)
   ❑ Myself ❑ Friends ❑ Family ❑ Park management agency ❑ Public ❑ Other (specify)__________

10. Please state whether you **DISAGREE** or **AGREE** with the following statements concerning your visit.

Visits to this park help me to:	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Better understand nature	1	2	3	4	5
Enjoy attractive natural scenery	1	2	3	4	5
Interact with people	1	2	3	4	5
Build and strengthen my relationships with others	1	2	3	4	5
Improve my physical health	1	2	3	4	5
Improve my mental health	1	2	3	4	5

 Continue on the next page

11. How **IMPORTANT** are the following factors to you during your visit to the park? (Circle **ONE** for each item)

	Not at all Important	Slightly Important	Moderately Important	Very important	Extremely important
Having some time with my family/friends	1	2	3	4	5
Being with and / or observing others in the area	1	2	3	4	5
Experiencing calm, peace, and solitude	1	2	3	4	5
Releasing tension and escaping from daily routine	1	2	3	4	5
Enjoying nature	1	2	3	4	5
Learning or experiencing something new	1	2	3	4	5
Getting exercise	1	2	3	4	5
Being physically healthy	1	2	3	4	5
Going other places easily without crossing a road	1	2	3	4	5

12. How satisfied are you with the following aspects of the park? (Circle **ONE** for each item)

	Not Satisfied	Somewhat Not Satisfied	Neutral	Somewhat Satisfied	Satisfied
Easy access to the park (parking, etc.)	1	2	3	4	5
Connections to attractions/other destinations	1	2	3	4	5
Overlooking scenery /observation decks	1	2	3	4	5
Trees and vegetation	1	2	3	4	5
Historical elements associated with the railroad	1	2	3	4	5
Facilities (restrooms, drinking fountain, etc.)	1	2	3	4	5
Places to sit and gather along the trail	1	2	3	4	5
Overall satisfaction with the park	1	2	3	4	5

13. Would you like to have this kind of park in your neighborhood? ❑ Yes ❑ No

13a. If **YES**, please check **ALL** reasons you want this park in your neighborhood.
❑ Increase economic growth
❑ Increase recreation opportunities
❑ Increase community pride
❑ A place to enjoy nature
❑ A place to rest and relax
❑ A place to interact with neighbors
❑ A place to get places without crossing a road
❑ Other (specify)______________

14. What social media accounts do you have? (check **ALL** that apply)
❑ Facebook ❑ Instagram ❑ Twitter ❑ Snapchat ❑ Other (specify)_______________

14a. How often do you look at the social media? ❑ Never ❑ Rarely ❑ Monthly ❑ Weekly ❑ Daily

14b. How often do you post the social media? ❑ Never ❑ Rarely ❑ Monthly ❑ Weekly ❑ Daily

15. What is your gender? ❑ Male ❑ Female

16. What is your age group? ❑ 18 - 29 ❑ 30 – 39 ❑ 40 – 49 ❑ 50 – 59 ❑ 60 – 69 ❑ over 70

17. What is your race/ethnicity? (check **ALL** that apply, if you prefer **NOT** to answer, please **SKIP** this question)
❑ White
❑ Black or African American
❑ Hispanic / Latino
❑ Asian
❑ American Indian
❑ Other (specify)____________

18. What is the highest level of education you have completed?
❑ Some High school
❑ High school/GED
❑ Bachelor's/Associate's degree
❑ Graduate/professional degree
❑ Postgraduate degree
❑ Other (specify)____________

19. Please indicate your total household income range before taxes last year.
❑ $25,000 or less
❑ $25,001 to 50,000
❑ $50,001 to 75,000
❑ $75,001 to 100,000
❑ $100,001 or 150,000
❑ $150,001 or over

20. Please write the ZIP code for your permanent address. ZIP code:______________

This is the end of the survey. Thanks for your time.

Appendix D: How to Get Insights from Social Media

Text mining, also known as text analytics, allows researchers to reveal the knowledge behind textual data. This method uses a set of fields, such as information retrieval, data mining, statistics, and linguistics, to extract meaningful information from unstructured textual data (He, Zha, & Li, 2013). The emergence of social media applications has contributed to the growth of text mining. Text mining borrows techniques from big data analytics such as natural language processing (NLP), machine learning (ML), information retrieval (IR), and data mining (DM).

Hu and Liu (2012) demonstrated the text mining process in three phases: text pre-processing, text representation, and knowledge discovery. Text pre-processing, also known as data processing, makes textual data more consistent for analysis (Table 22). Their study included stopword removal and stemming as text pre-processing tools. Stopword removal deletes all stopwords, which are considered meaningless in textual data. Stemming turns words into their stems on the basis of their root forms (Hu & Liu, 2012) to reduce confusion. Other tools for text pre-processing are lowercasing, normalization, and noise removal. Lowercasing of capital letters prevents the sparsity problem by ensuring that words differing only in case are mapped to the same form (Ganesan, 2019). Normalization transforms text into a standard form, for example by mapping both “stop-words” and “stop words” to “stopwords,” and standardizes spelling, for example by changing “wooooow” and “wuw” to “wow.” Text pre-processing can differ by research design and purpose; to analyze tweets, most researchers eliminate hashtags, URLs, HTML code, and ID before conducting text analytics.

Table 5. Examples of text pre-processing

	Input	Output
Stopword removal	***I*** walk ***on the*** High Line	walk High Line
Lowercasing	I walk on the High Line	i walk on the high line
	I WALK on the HIGH LINE	i walk on the high line
	I walk oN the HiGh LiNE	i walk on the high line

Stemming	I ***walk*** on the High Line I am ***walking*** on the High Line I ***walked*** on the High Line	I ***walk*** on the High Line I am ***walk*** on the High Line I ***walk*** on the High Line
Normalization	I walk on the Hiigh Line I welk on the High Line	I walk on the High Line I walk on the High Line

Table 6. Example of BOW

sentence	Bag of words (BOW)
(1) I walk on the High Line	“I,” “walk,” “on,” “the,” “High,” Line”
(2) and you also walk on the High Line	“and,” “you,” “also,” “walk,” “on,” “the,” “High,” “Line”

The next processing step is text representation. This uses a model known as “bag of words” (BOW), which is the most popular method for identifying and categorizing content. The basic procedure is to divide texts up into words and count each word’s frequency to visualize it (Zhang, Jin, & Zhou, 2010). A clustering algorithm is widely used to visualize the words. In the BOW model, a sentence or paragraph is the bag of its words.

Table 7. Example of BOW and TF.

Word in BOW	Term Frequency (TF)	The presence of words in: “I walk on the High Line”
walk	2	1 (yes)
on	2	1
the	2	1
High	2	1
Line	2	1
I	1	1
and	1	0 (no)
you	1	0
also	1	0

For example, the sentence “I walk on the High Line, and you also walk on the High Line” contains these unique words in the bag: “I,” “walk,” “on,” “the,” “High,” Line,” “and,” “you,” “also,” (Table 23). Each of these word has a number of occurrences in the sentence. This model can be used to calculate the number of occurrences of words to identify the characteristics of given textual information. After transforming textual information into a BOW list, researchers can measure term frequency or calculate Boolean variables based on the presence of words in the source (Table 24). TF-IDF weight is the most popular schema for measuring the importance of words (Hu and Liu 2012). Term frequency (TF) is a score based on a word's occurrences in the text, and inverse document frequency (IDF) is a scoring for the scarcity of each word. Thus:

TF(w) = Number of times w appears in the data/total number of words in the data.
IDF(w) = $\log_e$(total number of data/number of data containing w).
TF-IDF = TF × IDF

The BOW model is helpful for identifying topics and activities in urban green spaces. Lim, Lee, Kendal, Rashidi, Naghizade, Feng, and Wang (2019) examined park visitors' sentiments and activities using text analytics (Lim et al., 2019). In this study, we use latent Dirichlet allocation (LCA) to detect topics and activities in urban green spaces. LCA represents a topic or activity as a bag of words for clustering purposes. Two main findings arise: The identification of common activities in large parks, such as eating, drinking, and events; the other is that park visitors like to take photos at landmarks and historic sites. These are predictable results that are mentioned in many studies (Sather-Wagstaff, 2008). However, the value of this study is in examining the use of text analytics to detect the same behaviors as are found using traditional methods such as surveys.

One approach using the BOW model is the N-gram language model, which is widely used in text mining to derive meaningful implications from textual information. An N-gram is a sequence of N words occurring together in a given set of textual data (Broder, Glassman, Manasse, & Zweig, 1997) For example, consider the sentence "I walk on the High Line." When N = 2 (a bigram), the sentence yields "I walk," "walk in," "in the," "the High," and "High Line." When N = 3 (a trigram), the results are "I walk in," "walk in the," "in the High," and "the High

Line." N-grams are widely used in statistical NLP to calculate the joint probability of a sentence (Martin & Jurafsky, 2009). To compute vector scores and probabilities of sequences, many studies use the N-gram model when detecting issues (Laniado & Mika, 2010), behaviors (Myslín, Zhu, Chapman, & Conway, 2013), sentiments (Tang et al., 2014), and social unrest (Compton, Lee, Lu, Silva, & Macy, 2013). Agrawal, Budak, El Abbadi, Georgiou, and Yan (2014) used the term-frequency model to monitor a subset of stream elements (Agrawal, Budak, El Abbadi, Georgiou, & Yan, 2014).

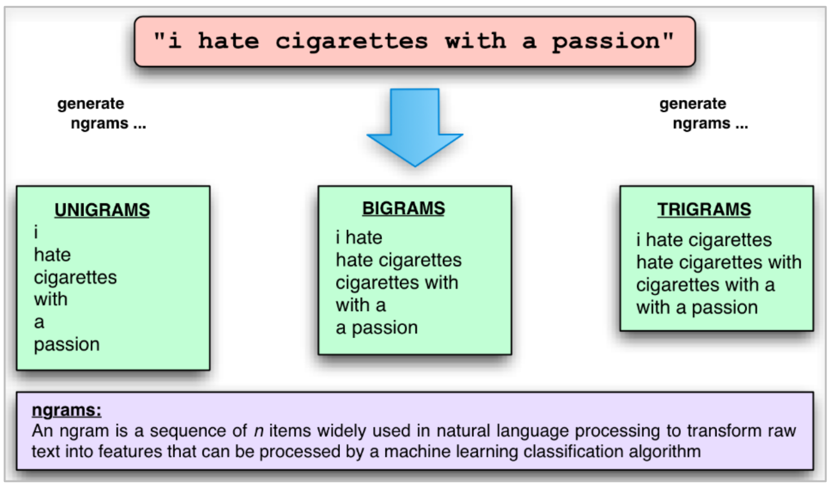


Figure 29. Examples of N-grams (Myslín et al., 2013).

In the case of detecting behavior, Myslin, Zhu, Chapman, and Conway conducted a study to understand behaviors and attitudes to tobacco and develop a machine-learning classifier to detect them (Myslín et al., 2013). They collected more than 7,000 tweets related to tobacco from 2011 to 2012 and manually classified them. Using these data, they trained classifiers to detect the same sentiments (Myslín et al., 2013), and a bigram model was used to calculate the probability of sequential words for the classification algorithm (Figure 29).

In the final processing for text analysis, several tools are used, such as classification, clustering, event detection, and sentiment analysis (Hu & Liu, 2012). Sentiment analysis is generally used in text analytics, especially with social media data. According to Mukherjee, sentiment analysis is a “natural language processing and information extraction” task to understand user’s feelings from positive or negative tweets by analyzing the text. It can be considered the study of people’s opinions, sentiments, emotions, and attitudes (B. Liu, 2015). Among the sentiment analysis classifications, the lexicon-based approach is used in the present study. This approach identifies positive and negative sentiment terms by using a dictionary (Maynard & Funk, 2012), collecting a small set of opinion words, and then growing this set by searching a large collection of text (Parrott, 2001).

There are two ways to detect sentiment in textual information: polarity analysis and emotion analysis. Polarity analysis sorts sentiments into three kinds: negative, positive, and neutral. Emotion analysis categorizes feelings as happy, sad, fearful, angry, surprised, and disgusted (Zhai & Massung, 2016). Although sentiment analysis is helpful for understanding human interests, the inherent nature of social media data poses challenges to its validity because of the shortness of the postings and the tendency to use abbreviations (Hutto & Gilbert, 2014). To reduce these difficulties and improve the method’s validity, Hutto and Gilbert (2014) provided a model called the valence aware dictionary for sentiment reasoning (VADER). They examined the validity of a sentiment lexicon and then combined lexical features with five rules embodying words. Then they examined the model’s accuracy with social media data by comparing its results with the facts. The authors claimed the following strengths for VADER : (1) It works well on short and abbreviated texts. (2) It is constructed from a human-created sentiment lexicon. (3) It is fast enough to analyze streaming data. (4) It does not differ from a speed–performance tradeoff (Hutto & Gilbert, 2014).

They then developed a new sentiment lexicon list from existing sentiment lists, such as Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001), Affective Norms for

English Words, and General Inquirer, and added human-created lists. To make these they used a wisdom-of-the-cloud (WotC) method to estimate a point for the sentiment of each feature. Each feature given a score by a human, from –4, extremely negative, to +4, extremely positive, with 0 as neutral. In this way, they identified 7,500 lexical features with validated scores indicating their sentiment polarities. Using these lexicons, VADER measures each word's sentiment score, sums them, and normalizes the final score to between –1 and +1. Positive sentiments is then treated as a final scores of >0.05, negative sentiment <0.05, and neutral as between those. Unlike other sentiment-analysis lexicons, VADER represents human sentiment, especially in tweets (Hutto & Gilbert, 2014).

Since Hutto and Gilbert introduced VADER in 2014, many studies have used it for text analysis of tweets. Becken, Stantic, Chen, Alaei, and Connolly (2017) conducted a study monitoring human sentiment about the Great Barrier Reef system in Australia. They collected geo-tagged tweets posted from the coral reef system to identify their contents and distribution, in order to examine the possibility of using social media in environment studies. Then they conducted a sentiment analysis using the VADER lexicon. They concluded that tweets provide pools of information covering large areas, and that the VADER lexicon is a useful tool for Twitter data (Becken, Stantic, Chen, Alaei, & Connolly, 2017).

In the case of landscape architecture, Brindles, Cameron, Ersoy, Jorgensen, and Maheswaran (2019) conducted a sentiment analysis using the VADER lexicon to compare green space quality and self-reported health with survey and social media data, particularly from Flickr. Their results showed a poor correlation between social media and survey quality about green space quality (Brindley, Cameron, Ersoy, Jorgensen, & Maheswaran, 2019). Although they did not recommend the use social media for assessing green space quality, their study did contribute to the literature as the first attempt to examine social media data as a proxy.

Strengell and Sigg (2018) also conducted a sentiment analysis using the VADER lexicon to understand human–environment interactions in urban areas. They suggested social media data as complement to surveys that are unable to detect real-time results. They collected real-time tweets about urban areas in Helsinki through a streaming API and then mapped each post's sentiment score to a heat map. Because of the limits of translation, the VADER toolkit showed only 96% accuracy in classification. However, the authors were able to visualize citizens' sentiments and behaviors on the map in real time (Strengell & Sigg, 2018). Those studies support the use of the VADER lexicon for analyzing Twitter data.

Appendix E: Terminology

Big data

The arrival of big data marks a significant change in data collection, storage, and analysis. Provost and Fawcett (2013) defined big data as data sets that are too large for traditional data-processing systems and that therefore require new technologies. According to Zikopoulos, deRoos, Parasuraman, Deutsch, Giles, and Corrigan (2013), big data have four dimensions: (1) volume, (2) velocity, (3) variety, and (4) veracity. In this study, “big data” refers to large, real-time, user-generated data sources.

Small Data

There are several ways to understand how people use places and interact with each other in them. Researchers have developed several ways to measure these variables through field observation, surveys, and interviews. Because those methods have been used in the study of urban parks for a long time, they are often called “traditional analytics” (Aoki, 1991; Eiter & Vik, 2015). Some researchers also call them “small data analytics,” in comparison with newer “big data” methods (Carr & Lassiter, 2017). In this study, we call the methods traditional to stress their historical status.

Social Media

Kaplan and Haenlein (2010) classified social media into several categories: collaborative media, such as Wikipedia; systems designed to share content, such as YouTube and Flickr; social networking platforms such as Twitter and Facebook; and blogging platforms, such as WordPress and Tumblr (A. M. Kaplan & Haenlein, 2010).

Users

In this study, “users” refers to social media users who may or may not post about elevated parks.

Visitors

"Visitors" refers to people who visited elevated parks and participated in a survey.

Travel Pattern

Travel patterns are used to understand how often people visit linear parks and how they do so. Detailed information about travel patterns of visitors can be divided into five parts: (1) number of visits (including the present one), (2) time spent in the park, (3) mode of visiting, (4) actual distance traveled, and (5) perceived distance traveled. Time spent in the park was asked in ordinal answers to simplify answers. For mode of visiting, respondents were asked to check all modes that applied. For perceived distance, respondents were asked about the difficulty of their travel on a 5-point scale (1: difficult to 5: easy).

Activity

Visitors were asked about all the activities they participated in in the park. These were grouped into five categories: physical, social, viewing, relaxing, and no main activity. These categories will be tested again for their reliability using commonality and factor analysis. Physical activities include biking, hiking, walking, jogging, running, and walking a dog. "Social activity" means posting to social media, as a virtual activity to enhance social cohesion. Viewing activities include all observing and sightseeing, such as looking from an observatory and viewing public art and wildlife, such as plants, flowers, and birds. Relaxation and no main activity included all forms of relaxation.

Motivation

Motivation is an important factor in park visits. Visitors were asked to rate the importance of each of nine motivational factors in their visit, including social, relaxation, contact with nature, physical health, and connection, on a 5-point Likert scale (1: not at all important, 2: slightly important, 3: moderately important, 4: very important, 5: extremely important). Two factors related to connection and learning things were excluded after commonality and factor analysis.

Perceived Benefits

Perceived benefits of park visits were divided into three kinds, derived from commonality and factor analysis: improvements to health, being in contact with nature, and improvements to social relationships. Those three were included in the survey to examine the benefits of linear parks in general. Respondents rated each on a 5-point Likert scale (1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree).

Satisfaction

Respondents were asked about their satisfaction with their overall experience, and with the connectivity and uniqueness of the park. Connectivity and uniqueness were chosen as the main characteristics of linear parks. All three factors were rated on a 5-point Likert scale (1: not satisfied, 2: somewhat unsatisfied, 3: neutral, 4: somewhat satisfied, 5: satisfied).

Social media usage

Social media usage was included in the survey to learn how park visitors used social media platforms, not only in linear parks but in their daily lives. Users were asked how often they looked at social media and how often they posted on social media platforms. To measure their social media usage in parks, the survey asked visitors whether they posted their experiences in parks, and if so the reason for their posting and who they wanted to share it with.

Appendix F: Three Sites

1. The 606

The 606 was formerly the Bloomingdale Line, a linear railroad crossing Chicago from east to west. The railroad was constructed in 1873 and was raised in the 1910s to reduce pedestrian fatalities. It was used for passenger and freight trains, but stopped operations in 2001. From 2002 to 2004, the greenway concept was introduced to the public, and Friends of the Bloomingdale Trail was launched in 2003 to advocate the use of the railway as a park (Gobster et al., 2017). In 2015, the 606 opened as the longest linear park in the U.S.

After the Great Chicago Fire, the city needed a railroad for rebuilding, and the Chicago and Pacific Railroad Company built this railroad in the northwestern part of the city (The 606, 2019). It connected passengers, delivered goods, and supported industrial growth across the city. Many new people moved to Chicago, and the population increased dramatically, especially in the areas adjacent to the railroads (Sawislak, 1995). But thousands of people were injured or killed every year because the railroads were on the ground level. As a result, in 1893 the city approved elevating the railroads, and the Chicago and Pacific Railroad Company built raised tracks. This affected lot sales and increased the construction of residential areas (Fellmann, 1957). From the 1910s to the 1990s, this railway delivered products and passengers (Rail-Trail History, 2019).

In 1997, the Chicago Metropolitan Agency for Planning included the Bloomingdale Line in its Bicycle Facilities Development Plan (City of Chicago). In 2003, the city announced a plan to convert the railroad (The 606, 2019), and in 2004, it included this project in the Chicago Park District's Logan Square Open Space Plan (LSOSP) for turning unused railroads into green spaces. The LSOSP aimed to (1) increase public open spaces such as parks, plazas, greenways, and outdoor places, (2) improve the quality of existing open spaces, (3) investigate recreational and open spaces on the Bloomingdale rail line, and (4) work with communities. The initial plan for the 606 was to provide boulevards, bike routes, and trails on the railroads (City of Chicago, 2004).

In 2007, the Chicago Architectural Club opened an exhibition of 26 conceptual designs exploring the conversion of the Bloomingdale Line. These proposals envisioned the future use of railroads across the city as ecological infrastructure, flowering gardens, green viaducts, and greenhouses (Sinha, 2014). One proposals was the Bloomingdale Park and Trail, by Michael Van Valkenburg Associates (MVVA). They proposed turning the railroads into walking paths and bike trails for residents. After that, MVVA developed the conceptual design, and the city selected the teams, ARUP, Ross Barney Architects, and MVVA in 2009. In 2013, the Bloomingdale Trail Project changed its name to “The 606,” referring to the community’s zip code prefix (Gobster et al., 2017). The 606 represents the essential goal of the project to connect communities. In 2015, the 606 opened to the public.

The design objectivities and features of the 606 are as follows: (1) enhancing the unique attributes of the Bloomingdale rail line, (2) balancing trail and park, (3) considering residential areas’ privacy, (4) providing integrated accessibility for all kinds of transportation in the city, and (5) allowing spaces for the arts (the City of Chicago). The 606 tries to preserve and capitalize on the feel of the Bloomingdale rail line by using the existing massive concrete structure filled with soil and restoring the industrial infrastructure. The design team focused on these features and maintained the structure to reflect the history of the rail line (City of Chicago, 2004). To balance trail and park, the 606 has separate bike and walking paths, shared-use paths, and observation spots with seating. To ensure safety, the designers were careful to add physical barriers. And because the 606 cuts across residential areas, separation between public and private spaces is an important issue; the designers addressed this with planting and the park's linear structure. To improve accessibility, the 606 has thirteen access points. The designers also considered the network of existing transportation and selected access points accordingly.

2. The High Line

The City of New York allowed the construction of street-level railroad tracks along Tenth and Eleventh Avenues on Manhattan's West Side in 1847 to ship commodities such as dairy and beef into lower Manhattan. Safety problems arose, and the area became known as "Death Avenue" (Friends of High Line, 2019). Thus in 1929, the city, the state, and New York Central undertook the West Side Improvement Project led by Robert Moses (Bighorse, 2010), and the tracks elevated. Because the line ran through the centers of several blocks, 640 buildings were demolished (Friends of High Line, 2019), and the construction split many communities in two.

The invention of the standard container cargo fostered the growth of the interstate trucking system during the 1950s (Waldheim, 2016) and led to reductions in rail traffic throughout the U.S. Thus in 1980, owner Conrail decided to disconnect the High Line railway from the national rail system for a year. But when it was reconnected in 1981, there was no demand, and the tracks went out of use. In 1999, the Friends of the High Line was formed by Joshua David and Robert Hammond with residents of nearby communities to advocate for the line's preservation and repurposing as a linear park. In 2004, the city government provided $50 million, and other funders raised the last $150 million needed for the park. In 2005, the federal Surface Transportation Board issued a certificate of trail use to allow the city to remove the line from the national rail system.

The High Line's neighborhood, Chelsea, was maintained as a diverse neighborhood until the 1980s. Although the area was underdeveloped because of the meatpacking district, this also led to a lot of small and diverse businesses, such as gay bars, art galleries, and grunge culture shops. Although the disused High Line became a place for drug transactions or and homeless gatherings, Chelsea retained its own characteristics based on its diverse culture (David & Hammond, 2011).

David divided the neighborhoods adjacent to the High Line along five boundaries based on blocks. The first boundary, on the north side, was the 30th Street Railroad, which was owned by state-controlled agencies (David, 2002). This area includes a railyard and several old building parcels, which were considered unpleasant units. Below this, the blocks between 30th and 26th Streets were used for manufacturing and light industry. The next group, from 20th to 26th Street, are surrounded by art galleries, museums, and performance spaces and are considered an international art hub. The blocks between 20th and 14th Streets are transitional blocks between the art hub to the north side and the Meat Packing District to the south. This area was used for single houses and small galleries. The last boundary marked the most interesting area: the Gansevoort Meat Packing District. This area represents much of the historical character of New York City, which offered great potential to the High Line.

Friends of the High Line conducted a design competition in 2003. This was open to everyone and aimed not at developing design concepts, but also at promoting public interest in the High Line. A total of 720 ideas were collected and exhibited at Grand Central Station (David & Hammond, 2011). In 2004, the Friends of High Line looked for something buildable in the professional competition. Fifty-two proposals arrived at the office, and a team led by Field Operation won the competition. As with the idea competition, the selected projects were exhibited at the Museum of Modern Art. Hammond reported that it “didn’t change anything about the legal, political, or financial hurdles that lay ahead. But once we were at MoMA, people thought the High Line was going to happen" (Friends of the High Line, 2019).

The main concept that penetrated the winner's design was melancholia. The winning team focused on the "placeness" of the High Line: dark and mysterious, with a sense of melancholy and otherworldliness. The views represented this well.

The main design strategies had three features: (1) varied proportions between paving and vegetation, (2) the vegetation itself, and (3) memorializing the design of the railway. Field Operations provided several prototypes of paving along the High Line, ranging from 100% paving to 0% paving (i.e., 100% vegetation). Because the High Line is long and narrow, if the same paving patterns are repeated, users may feel bored. This strategy creates opportunities for users to have varied experiences, and it helps the designers overcome the restrictions of linear parks, such as limited soil depth and the dryness of the surface (Fehrenbacher, 2014).

The second feature in the design was the vegetation. The planting design of the High Line aimed to retain the existing bio-diversity and ensure sustainability by using the self-seeded landscape. In the 52 years since operations had ceased, native vegetation had overgrown the line. Piet Oudolf, the garden designer who designed the High Line, described its planting scheme as inspired by many natural processes and as creating different moods to evoke the patterns of nature (Friends of High Line, 2019).

The third feature was the representation of the history of the tracks. To memorialize the railway, a couple of strategies were employed, such as peel-up design language on benches, bike racks, planters, and a fountain, ash paving patterns representing the old rail tracks. Some sections of actual track also remain to let visitors know where they are.

Finally, there are many overlook points on the High Line. From its endpoint in the meatpacking area to the theatre, the High Line offers numerous places for visitors to enjoy views of New York City. Overall, the design of the High Line provides a diversity of experiences.

3. The High Bridge

The High Bridge is a part of the High Bridge Trail State Park in Virginia. This is a 31-mile trail on a railroad that was part of the Norfolk Southern Railway, which had 19,420 miles of routes in total (Norfolk Southern, 2018). After Norfolk Southern's last train stopped in 2004, the company donated 31 miles of railroad to the Virginia Department of Conservation and Recreation (DCR), who added the trail to the Virginia State Park system. In 2012, the DCR opened the park to the public for recreation (Virginia Department of Conservation and Recreation, 2019)

The High Bridge is the centerpiece of the park and is on the National Register of Historic Places for its history. The bridge itself was constructed in 1854 by Southside Railroad to connect Petersburg and Lynchburg across the Appomattox River (Revolvy, 2019). It is 2,400 feet long, and its height ranges from 60 to 125 feet. The bridge was originally made of wood tresses and had two levels: trains and pedestrians used the upper one, and wagons used the lower one. After the Battle of the High Bridge in April 1865, Confederate forces destroyed the bridge to block Union troops. Later, Robert Lee ordered the bridge rebuilt, and the trains resumed in September 1865 (Flippen & McClintock, 2014).

The High Bridge Trail State Park opened in 2012. The DCR set its objectives for the park in 2006, however. According to its report, the purpose of the park was (1) to afford non-motorized pathways for pedestrians, cyclists, and horseback riders, (2) to protect the historic and natural resources along the trail, and (3) to provide access to the region's historic events to individuals, communities, and the whole country. In 2012, the DCR identified four experiences of its visitors: rural remote, rural social, urban social, and focal point. The focal point of the trail, the High Bridge, is the highest point on the trail (DCR Summary, 2012).