Journal of Environmental Management 317 (2022) 115410

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Assessing the validity of mobile device data for estimating visitor demographics and visitation patterns in Yellowstone National Park

Yun Liang [a], Junjun Yin [b], Bing Pan [a,*], Michael S. Lin [c], Lauren Miller [d], B. Derrick Taff [a], Guangqing Chi [e]

[a] Department of Recreation, Park, and Tourism Management, Pennsylvania State University, State College, PA, 16801, USA
[b] Social Science Research Institute, Pennsylvania State University, State College, PA, 16801, USA
[c] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Kowloon, China
[d] Yellowstone Center for Resource, Yellowstone National Park, WY, 82190, USA
[e] Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, State College, PA, 16801, USA

ARTICLE INFO

Keywords:
Mobile device data
Visitor demographics
Temporal visitation patterns
National park

ABSTRACT

Monitoring visitor demographics and temporal visitation patterns can help national park managers understand their visitors and allocate resources more effectively. Traditional approaches, such as visitor surveys or vehicle counts, are limited by time, space, labor, and financial resources. More recently, mobile device data have been adopted for monitoring visitors in park-related or tourism research. However, few studies validated mobile device data with traditional visitor surveys or count data. Combining mobile device data with the American Community Survey (ACS), this study assessed mobile device data's validity in a national park context with three approaches: Points of Interest (POIs), visitor demographics, and temporal visitation patterns. The results revealed that only half of the POIs inside Yellowstone National Park are valid. Compared to traditional visitor surveys, mobile device data are limited due to platform bias and the exclusion of international visitors, resulting in discrepancies in visitor demographics, such as education and income levels. Conversely, mobile device data have strong correlations with count data regarding monthly and daily visitation patterns. The results suggest that with careful consideration, mobile device data can serve as an additional and complementary source of information to traditional survey data for understanding visitor demographics and temporal visitation patterns.

Credit author statement

Yun Liang: Conceptualization, Methodology, Data Formal analysis, Writing – original draft; **Junjun Yin**: Conceptualization, Investigation, Methodology, Writing – review & editing; **Bing Pan**: Conceptualization, Methodology, Investigation, Writing – review & editing; **Michael Lin**: Writing – review & editing; **Lauren Miller**: Data collection, Writing – review & editing; **Derrick Taff**: Writing – review & editing; **Guangqing Chi**: Conceptualization, Writing – review & editing

1. Introduction

Prior to the COVID-19 pandemic, the United States (U.S.) national parks attracted 327.5 million domestic and international recreation visits in 2019, which marked the fifth consecutive year of more than 300 million visitors (National Park Service Office of Communications, 2020a). Monitoring national park visitors and visitation numbers is essential for managing visitor flows, allocating resources, developing infrastructure, and predicting activity demands (Cessford and Muhar, 2003; English and Bowker, 2018; Pettebone and Meldrum, 2018; Rice et al., 2019). These data are vital to the mission of the National Park Service, enhancing visitor experiences and protecting park resources. For example, Ziesler and Pettebone (2018) indicated that visitor use data benefit the design of facility construction, including roadways, parking lots, visitor centers, and restrooms. Traditional approaches for collecting these data rely on visitor surveys or visit counts, which are limited by time, space, labor, and financial resources (Di Minin et al., 2015; Sessions et al., 2016).

The rapid development of information and communication technology generates big data, possessing the “3 V” characteristics, namely

* Corresponding author. ,.
E-mail addresses: yjl5451@psu.edu (Y. Liang), jyin@psu.edu (J. Yin), bingpan@psu.edu (B. Pan), michael.lin@polyu.edu.hk (M.S. Lin), lauren_miller@nps.gov (L. Miller), bdt3@psu.edu (B.D. Taff), gfc5047@psu.edu (G. Chi).

https://doi.org/10.1016/j.jenvman.2022.115410
Received 22 December 2021; Received in revised form 23 May 2022; Accepted 24 May 2022
Available online 10 June 2022

high volume, high value, and high velocity (Laney, 2001). Location-based Service (LBS) data from mobile devices, as one type of big data sources, has been adopted to investigate human mobility (Lee et al., 2020), the impacts of social distancing on economic inequality (Chiou and Tucker, 2020), and mobility and social networks in the U.S. (Chang et al., 2021). Internationally, researchers have explored factors influencing human mobility (Phithakkitnukoon et al., 2012), mobility and socioeconomic indicators (Pappalardo et al., 2015), mobility and event detection (Traag et al., 2011), and the representativeness issues of sparse mobile location data in Portugal, France, and China (Lu et al., 2017).

Mobile device data have also been applied in tourism-related research (Kubo et al., 2020; Park et al., 2020; Raun et al., 2016; Rodríguez et al., 2018). For example, Ma and Kirilenko (2021) compared social media data, mobile device data, and traditional surveys for estimating tourists' residency. By utilizing mobile positioning data, Raun et al. (2016) demonstrated that mobile device data could provide rich information on foreign visitors' characteristics and behaviors in Estonia in three dimensions, including spatial, temporal, and compositional.

National parks are diverse with vast expanses of landscape, providing a variety of visitor experiences. This makes visitor monitoring challenging. Mobile device data has significant potential as a complementary or alternative data source to traditional visitor survey/count data for studying park visitor behaviors and characteristics. Thus, this type of assessment on the validity of mobile device data is necessary and a prerequisite for its further applications (Ma and Kirilenko, 2021; Monz et al., 2019, 2021).

Although previous studies have investigated the effectiveness of using mobile device data to investigate human behaviors and its characteristics (Kang et al., 2020; Pappalardo et al., 2015; Traag et al., 2011; Ma and Kirilenko, 2021; Monz et al., 2021) and adopted mobile device data at a global scale, to date few studies investigated the validity of mobile device data. Most related studies have focused on studying user movement patterns with mobile device data (Creany et al., 2021; Merrill et al., 2020), while a limited amount of research has considered the underlying user demographics (Ma and Kirilenko, 2021; Monz et al., 2021).

To fill this gap, this current study aims to validate mobile data by comparing them with traditional visitor surveys and count data in a national park context. Combined with American Community Survey (ACS), this study follows three validation approaches for validation: 1) Point-of-Interests (POIs) from mobile device data, 2) visitor demographics, and 3) temporal visitation patterns.

2. Related work

Traditional approaches to collecting visitor demographics rely on visitor surveys (Pettebone and Meldrum, 2018). For example, Yellowstone National Park Summer 2018 Visitor Use Survey examined visitors' age, education, gender, race, and household income (National Park Service, 2019). However, these collection methods require substantial time and financial investment (Leggett et al., 2017). Survey interception is limited by specific surveying periods and a few selected areas in a given park (Cessford and Muhar, 2003; Di Minin et al., 2015; Hadwen et al., 2007). Language issues and administrator bias may also pose obstacles to capturing a valid demographic representation (Gstaettner et al., 2020). Additionally, conducting onsite visitor surveys may be more challenging in the era of the COVID-19 pandemic, given that visitors may have safety concerns while engaging with surveyors.

In terms of assessing visitation volumes, national parks typically use individual counts or proxy counts, and the level of detail beyond entrance stations varies considerably by unit (Ziesler and Pettebone, 2018). For individual counting methods, sun reflections and refractions, wildlife interference and movement, and temperature can impact the accuracy of traditional automated counters. Given staffing limitations, these systems often lack calibrated correction factors (Pettebone et al., 2010). Beyond automated traffic and trail counters to gauge use (Lawson et al., 2011, 2017; Pettebone et al., 2013, 2019; Ziesler and Pettebone, 2018), park or transportation ticket sales are one of the proxy counting approaches. For example, counting vehicles entering the park is a common proxy count approach, and associated vehicle counts must be multiplied by a parameter of persons-per vehicle in order to obtain estimated visitor counts (Ziesler and Pettebone, 2018). In summary, each method listed above contains significant biases and limitations.

In the United States, 97% of adults own at least one mobile phone (Pew Research Center, 2021). Researchers have taken advantage of data generated by mobile devices to understand human mobility. Since 2007, mobile device data has been applied in park-related and tourism research, such as investigating spatial and temporal patterns (Juhasz and Hochmair, 2020), visitor flows (Kupfer et al., 2021), and origins of tourists (Ma and Kirilenko, 2021). The application of mobile device data allows researchers to explore visitor behaviors and characteristics across a longer time scale and assess spatial variations across a larger region compared to visitor surveys (Alba et al., 2022; Kupfer et al., 2021). Additionally, mobile device data minimize time-intensive fieldwork for researchers and park staff (Monz et al., 2021).

Various companies providing mobile device data include SafeGraph, StreetLightData, and UberMedia (now part of Near, Rice et al., 2022). For example, Monz et al. (2019) employed mobile device data from StreetLightData to estimate monthly visitation in protected areas in Orange County, California, and validate it with traditional count data. Creany et al. (2021) estimated trail use and spatial distribution of visitors with mobile device data in protected areas in the same county. Juhasz and Hochmair (2020) investigated temporal visitation patterns, distance from home, and event detections in three Florida cities using SafeGraph data. Kupfer et al. (2021) employed SafeGraph data to investigate the effects of the COVID-19 pandemic on changes in temporal visitation patterns and visitor flows in six U.S. national parks.

In summary, traditional studies on visitor demographics and temporal visitation patterns have often relied on visitor surveys and individual or proxy counts. New types of mobile phone data are now available and being adopted in tourism and park settings, potentially eliminating some of the limitations of small temporal and spatial scales of visitor surveys. However, limited attempts have been made to assess the validity of these types of mobile data. Therefore, this study intends to validate mobile device data in a national park context, combined with ACS, with three approaches, POIs, visitor demographics, and temporal visitation patterns.

3. Methodology

3.1. Data

Yellowstone National Park (YNP), one of the most visited national parks in the U.S. (National Park Service, 2022a), was selected as the context of this research for one reason: the authors have access to a visitor use survey, count data and mobile device data in YNP at the same time period.

Five datasets were utilized for this investigation (Table 1), including YNP Summer 2018 Visitor Use Surveys (2018), NPS Stats Recreation Visits by Month (2018–2020), Trails/Gates daily count data in Yellowstone National Park (2018 summer & 2019 summer), SafeGraph mobile device data (2018–2020), and ACS 2015–2019 (at Census Tract level).

3.1.1. YNP Summer 2018 Visitor Use Surveys

YNP Summer 2018 Visitor Use Surveys collected visitor demographics and visitor experiences at various attractions in YNP across the summer season of 2018 (National Park Service, 2019). Visitor demographics, such as age, gender, race, educational level, income level, and origins of residency, were retrieved from the survey report.

Table 1
Data sources.

Dataset	Source	Time Period	Accessibility
Yellowstone National Park Summer 2018 Visitor Use Surveys	NPS	2018	Public
NPS Stats Recreation Visits by Month	NPS IRMA	2018–2020	Public
Trails/Gates daily count data in Yellowstone National Park	Collected by NPS administration	2018 summer & 2019 summer	Private
SafeGraph Core Places/Patterns	SafeGraph	2018–2020	Public for academic researchers
American Community Survey	The United States Census Bureau	2015–2019 (5-Year Estimates)	Public

3.1.2. NPS Stats Recreation Visits by month

NPS Stats Recreation Visits by Month provided the number of recreational visitors by month from 1979 to the current calendar year for U. S. national parks (National Park Services Stats, 2021). The monthly data of YNP in the same time period was utilized for the validation of mobile device data.

3.1.3. Trails/gates daily count data

YNP management provided two datasets containing trail and entrance gate daily counts. The trail count dataset included visits in popular locations such as Old Faithful Spring, Lower Fall, and Artists Paintpots, covering June 2018 to November 2018 and June 2019 to November 2019. The entrance gate vehicle count dataset included data from May 2019 to October 2019 and May 2020 to October 2020. The two datasets were utilized to compare the daily totals with the visits in SafeGraph data. The trail counter data were calibrated for a number of hours for each season, and multipliers were determined after the compilation of all calibration data.

3.1.4. SafeGraph Data

SafeGraph is a commercial company that provides Point of Interest (POI) and Location-Based Services (LBS) data in the U.S, Canada, and the United Kingdom (Juhasz and Hochmair, 2020). POI refers to a specific useful or interesting location. The POI data are compiled from several sources, including mobile phone GPS data and open government data. SafeGraph can track anonymous locations from mobile applications after obtaining opt-in consent[1] from their users. These data do not contain any identifiable information, such as usernames or the MAC address of mobile devices. They only include the latitude and longitude of a device at a given location and time. SafeGraph employs this geographic location information to determine the number of visits to each POI (SafeGraph, 2021).

The main product of POI data is SafeGraph Places, consisting of three datasets: *Core Places*, *Patterns*, and *Geometry* (SafeGraph, 2020). The *Core Places* dataset provides POI data and related attributes for non-residential locations, including geospatial coordinates, addresses, brand affiliation, open hours, and locational categories (restaurants, national parks, museums, etc.). The current scope includes restaurants, general stores, malls, parks, hospitals, museums, offices, etc. (SafeGraph Places Manual, 2020).

SafeGraph compiles POI data in the following steps:, collecting public location data on the web, applying public APIs and collecting updated locations from open web domains, processing and modeling to infer additional attributes (e.g., inferring the category of a POI), and collaborating with a third-party on additional data sources to fill in gaps. Once SafeGraph compiles all data sources, they go through a strict cleaning and incorporating process to ensure the accuracy and currency of the dataset (Bonack, 2021).

SafeGraph *Patterns* provides LBS data in the format of visit counts to POIs. SafeGraph aggregates and anonymizes mobility data from mobile applications with which users have allowed the tracking of their locations. SafeGraph associates visit characteristics (e.g., daily raw visits) to specific places by utilizing its *Core Places* and *Geometry* datasets. The SafeGraph *Patterns* dataset also provides specific locations where people travel to and from. SafeGraph aggregates information at the Census Block Group (CBG) level of a device's home location (Bonack, 2021). Only a CBG with at least two devices is included (SafeGraph, 2020). To determine people's home CBG, SafeGraph analyzes six weeks of data during nighttime hours (treating 6 p.m. to 7 a.m. as common nighttime) and assigns a home location for a mobile device (SafeGraph, 2020). SafeGraph also includes a *Geometry* dataset that contains geographic boundaries of POIs formatted as Well-Known Text (WKT) (Bonack, 2021). SafeGraph infers the shape of POIs by integrating with reliable third-party satellite imagery and applying machine learning approaches.

SafeGraph data were utilized because of their free availability for academic researchers. We retrieved SafeGraph data from *Core Places* and *Patterns* datasets. The *Core Places* dataset involves about 8.4 million POIs and related information, such as geographic location (latitude & longitude), address, category, NAICS CODE,[2] open hours, brands, and unique SafeGraph IDs. The *Patterns* dataset contains POIs with unique SafeGraph IDs, raw visit counts (monthly), visits by day (daily), visitor origins, etc. Currently, SafeGraph *Patterns* provide visitation data from January 2018 to November 2020. The two datasets, *Core Places,* and *Patterns*, can be linked by the same SafeGraph ID.

3.1.5. American community survey

American Community Survey (ACS) is a demographics survey program conducted by the U.S. Census Bureau. First, it regularly collects information on American households on educational attainment, income, disability, employment, housing characteristics, etc. Next, the Census Bureau aggregates individual ACS responses into various grouped geographic levels. These levels are legal and administrative entities such as states, counties, cities, and congressional districts, as well as statistical entities such as metropolitan scale and tracts. ACS can be associated with SafeGraph data to extract visitor demographics at a County/Census Tract/Block Groups level. To align with the corresponding time period, this study utilized ACS 2015–2019 (5-Year Estimates) (United States Census Bureau, 2020b).

3.2. SafeGraph POI selection and validation

We followed three steps in order to select and examine SafeGraph POIs inside YNP. In the first step, we performed a geospatial operation, namely point in polygon, to locate those POIs with coordinates that fall within the boundary of YNP. As a result, 80 POIs from SafeGraph *Core Places* were found.

In the second step, for validation purposes, Google Maps was employed to assess the location names, geographic coordinates, and addresses for each POI. First, the location names were searched in Google Maps. If the given location name was found and appeared to be accurate, the geographic coordinates and exact addresses were retrieved from Google Maps. Next, the geographic coordinates and detailed addresses gained from Google Maps were compared to the latitude, longitude, and address provided by SafeGraph. As a result, 40 invalid POIs were filtered out among the original 80 due to unmatched geographic coordinates and/or inaccurate location names. Fig. 1 demonstrates the 40 valid POIs used in this study.

[1] Opt-in refers to companies explicitly asking users for permission to collect and process their personal data.

[2] Federal statistical agencies utilize NAICS as the standard to classify business establishments (U.S. Census Bureau, n.d.).

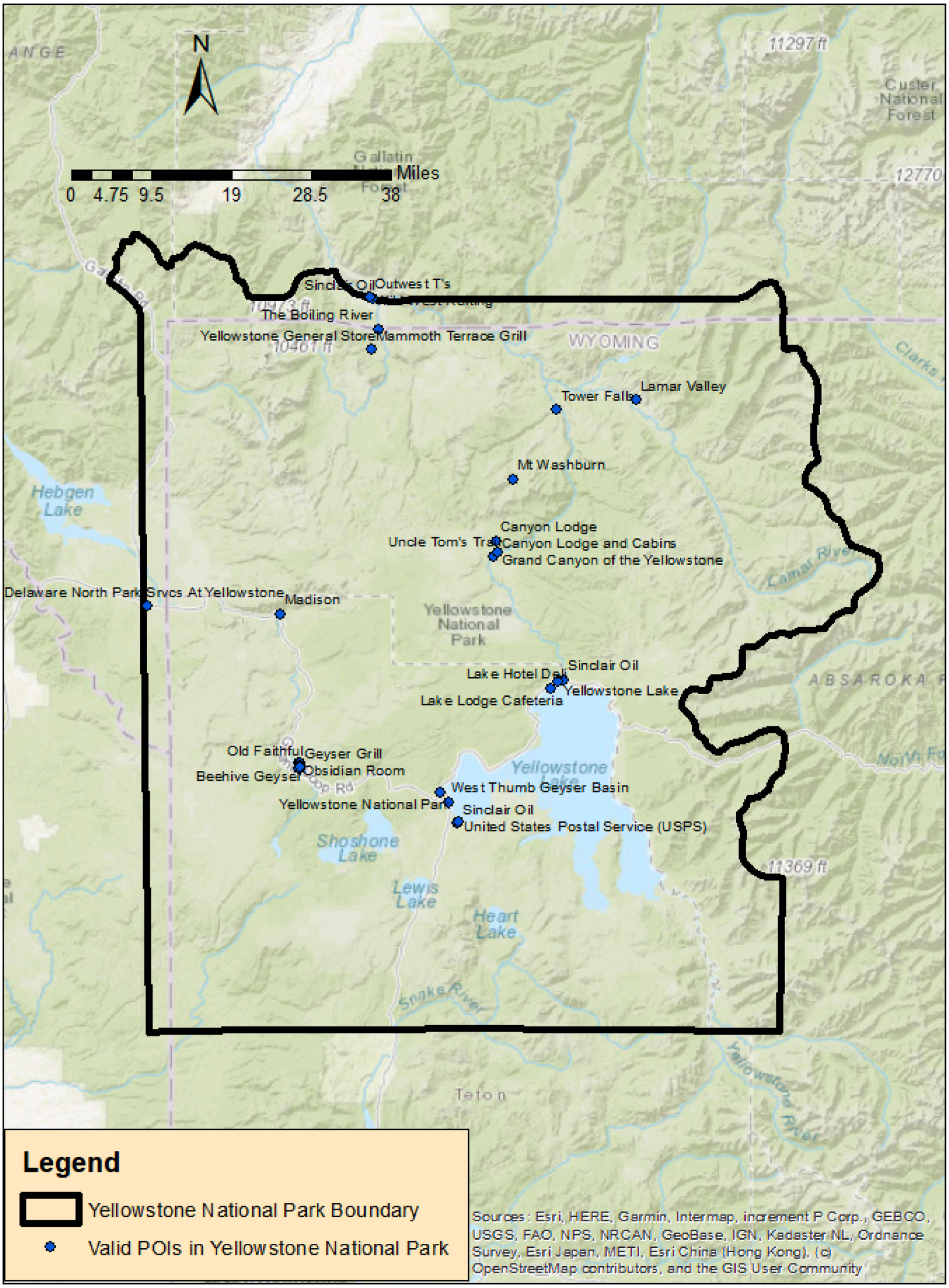

Fig. 1. Locations of valid POIs in yellowstone national park.

In the third step, the unique SafeGraph IDs of the 40 POIs from SafeGraph *Core Places* were utilized to retrieve the related visitation patterns from the SafeGraph *Patterns* dataset. Nine POIs (Table 2) were selected for further investigation. According to SafeGraph, these nine POIs represented the most popular attractions and possessed visitation numbers of more than 1,000 visits between May 2018 to September 2018. This period matches the data collection period of the YNP Summer 2018 Visitor Use Surveys.

Table 2
Selected top nine SafeGraph POIs for validating visitor demographics.

Location Name	Top Category	Latitude	Longitude	Street Address	City	Region	Postal Code
Yellowstone Art & Photography Center	Other Professional, Scientific, and Technical Services	44.459	−110.827	2 Old Faithful Rd	Wyoming	WY	82,190
Old Faithful General Store	Office Supplies, Stationery, and Gift Stores	44.457	−110.828	Old Faithful	Yellowstone National Park	WY	82,190
Canyon Lodge Cafeteria	Restaurants and Other Eating Places	44.734	−110.491	1 Grand Loop Rd Canyon Village Yellowstone National	Canyon	WY	82,190
Canyon Lodge and Cabins	Traveler Accommodation	44.734	−110.490	Canyon Village North Rim Dr	Yellowstone National Park	WY	82,190
Old Faithful Observation Point	Museums, Historical Sites, and Similar Institutions	44.460	−110.828	Old Faithful Village	Yellowstone National Park	WY	82,190
Obsidian Room	Restaurants and Other Eating Places	44.457	−110.830	Old Faithful Snow Lodge	Yellowstone National Park	WY	82,190
Geyser Grill	Restaurants and Other Eating Places	44.457	−110.829	1000 Old Faithful	Yellowstone National Park	WY	82,190
Fishing Bridge General Store	Office Supplies, Stationery, and Gift Stores	44.565	−110.375	1 N.E. Entrance	Yellowstone National Park	WY	82,190
Outwest T's	Clothing Stores	45.030	−110.707	228 Park St	Gardiner	MT	59,030

3.3. Data analysis methods

3.3.1. Visitor demographics

3.3.1.1. Merging SafeGraph with ACS. In the SafeGraph *Patterns* dataset, each POI is associated with visitation numbers and related visitor home CBGs. CBG is a geographical unit used by the U.S. Census Bureau and is the smallest geographical unit for which the bureau publishes related demographic data. Typically, Block Groups have a population of 600 to 3,000 people. A 12-digit identification number of each CBG has a GEOID[3] structure: 2 + 3+6 + 1 (12-digit number) = State + County + Tract + Block Group. The first two digits represent the visitor's home state; the first five (2 + 3) digits indicate the county; the first eleven (2 + 3+1) represent the Census Tract. Therefore, SafeGraph datasets can be associated with ACS and used to extract related demographic composition of residents at a State/County/Census Tract/CBG level.

Since the Visitor Use Surveys were conducted from May 2018 to September 2018, SafeGraph data from the same period were extracted. Combined with ACS results by CBGs, five demographic and socioeconomic variables, including gender, age, race, educational level, income level, and visitors' home states, were calculated from SafeGraph data. These five demographic variables and origins of visitors were selected in order to match with the Visitor Use Surveys.

3.3.1.2. Calculating visitor demographics. We employed two approaches to obtain visitor demographics from SafeGraph data: 1) the average visitor demographics of Yellowstone Arts & Photography Center (the most visited POI). This location became the most popular due to its proximity to Old Faithful Geyser, and SafeGraph did not capture the attraction as a POI; 2) the average visitor demographics of the nine most visited POIs (selected from the last step). We tested the second approach because applying only one POI to estimate visitor demographics may not represent the entire visitor population; the average of nine POIs can cover a broader range of visitors who may have only visited less popular POIs.

To calculate the aggregated visitor demographics, we adopted the following formula. Let X_t (t = 1, 2, …, n) be the percentage of each subgroup of each demographic variable in a Census Tract; let Y_t (t = 1, 2, …, n) be related visitation number in a Census Tract; the formula to calculate the percentage of each sub-group of each demographic variable of a POI is:

$$Percentage = \Sigma_{t=1}^{n} X_t Y_t \big/ \Sigma_{t=1}^{n} Y_t$$

However, the classification of subgroups for age, education level, and household income level are different between the visitor survey and ACS. To match the two data sources, we removed age subgroups below 18-year-old and recalculated the percentages of age subgroups 18-years-old or older to align with the Summer 2018 Visitor Use Surveys, which only collected data from visitors who were 18-year-old or older. For education levels, only one subgroup, namely less than high school, was kept for this study since the two data sources had very different definitions for other subgroups. Subgroups of household income levels in ACS were aggregated and recalculated to align with the subgroups in the survey.

After calculating the visitor demographics of each POI, the average visitor demographic of Yellowstone Arts & Photography Center and those of the nine POIs were compared with visitor demographics from the survey results by Chi-square Tests (National Park Service Stats, 2021). The top 10 states of visitor origins of the two data sources were presented in a descending order. Data aggregation and analyses for visitor demographics were conducted by 'pandas', a Python package for data analysis and manipulation.

3.3.2. Trends and correlations of monthly and daily visits

Before validating monthly and daily visitation patterns between the two datasets, SafeGraph data and the count data were normalized by the min-max feature scaling:

$$X' = X - X_{min} \big/ X_{max} - X_{min}$$

Next, SafeGraph data were compared with the count data from NPS Stats Recreation Visits (National Park Services Stats, 2021) and trail/-gate count data respectively regarding monthly and daily visitation patterns by Pearson's r correlation (Tenkanen et al., 2017).

Line charts of monthly and daily visit patterns of mobile device data and count data were created in Excel. Scatter plots with Pearson's correlations of the two types of data were created by 'ggplot2', an R package for data visualization.

4. Results

4.1. Visitor demographics and origins

Table 3 presents the average visitor demographics of the nine POIs and Yellowstone Arts & Photography Center and the results of Chi-square Tests with YNP Summer 2018 Visitor Use Surveys. This table also provides demographic composition at a national level.

[3] United States Census Bureau (2020a) defines that "GEOIDs are numeric codes that uniquely identify all administrative/legal and statistical geographic areas for which the Census Bureau tabulates data".

Table 3
Comparison of visitor demographics between 2018 SafeGraph data (combined with ACS) and Yellowstone National Park Summer 2018 Visitor Use Surveys.

Demographics	2018 Visitor Use Surveys (n=1,425)	2018 SafeGraph Nine POIs' Average (n=2,272)	Chi Square Statistics	P value	2018 Yellowstone Arts & Photography Center (n=4,105)	Chi Square Statistics	P value	National Statistics
Gender								
Male	51.0%	49.6%	0.66	0.42	49.6%	0.85	0.36	49.2%
Female	49.0%	50.3%			50.4%			50.8%
Age								
Under 5 Years	NA	NA	NA	NA	NA	NA	NA	NA
5 to 9 Years	NA	NA	NA	NA	NA	NA	NA	NA
10 to 14 Years	NA	NA	NA	NA	NA	NA	NA	NA
15 to 17 Years	NA	NA	NA	NA	NA	NA	NA	NA
18 to 24 Years	12.0%	10.7%	1.50	0.22	11.0%	1.035	0.31	12.1%
25 to 34 Years	21.0%	15.9%	15.49	<0.0001	16.3%	16.08	<0.0001	18.0%
35 to 44 Years	16.0%	16.6%	0.23	0.64	16.6%	0.27	0.60	16.3%
45 to 54 Years	17.0%	17.1%	0.012	0.91	17.1%	0.012	0.92	16.8%
55 to 64 Years	17.0%	17.9%	0.52	0.47	17.6%	0.27	0.60	16.7%
65 to 74 Years	14.0%	13.2%	0.52	0.47	12.8%	1.28	0.27	11.8%
75 or older	3.0%	8.6%	45.03	<0.0001	8.6%	49.57	<0.0001	8.4%
Race								
White	82.0%	82.5%	0.12	0.73	83.3%	1.18	0.28	72.5%
Black or African American	0.0%	4.8%	NA	NA	4.4%	NA	NA	12.7%
American Indian or Alaska Native	1.0%	0.9%	0.10	0.75	0.8%	0.40	0.53	0.9%
Asian	17.0%	5.2%	138.47	<0.0001	4.7%	139.59	<0.0001	5.5%
Native Hawaiian or Pacific Islander	0.0%	0.1%	NA	NA	0.1%	NA	NA	0.2%
Other Race	NA	2.2%	NA	NA	2.4%	NA	NA	4.9%
Two More Race	NA	2.9%	NA	NA	2.9%	NA	NA	3.3%
Highest level of former education								
Less than High School	1.0%	7.0%	71.05	<0.0001	7.4%	80.52	<0.0001	12.0%
High School Graduate	8.0%	Census Tract data and Survey data have different categories in Education Level			Census Tract data and Survey data have different categories in Education Level			
Some College	10.0%							
Bachelor's degree	37.0%							
Advanced degree	35.0%							
Vocational/Trade School	2.0%							
Two-year college degree	7.0%							
Household Income Level								
Less than $25,000	9.0%	14.2%	23.40	<0.0001	14.4%	27.42	<0.0001	19.3%
$25,000 to $49,999	12.0%	17.8%	22.29	<0.0001	18.4%	31.00	<0.0001	21.3%
$50,000 to $74,999	18.0%	16.6%	1.28	0.26	16.9%	0.95	0.33	17.2%
$75,000 to $99,999	17.0%	13.3%	9.50	0.0021	13.5%	10.44	0.0012	12.7%
$100,000 to $ 149,999	21.0%	17.8%	5.83	0.016	17.9%	6.59	0.010	15.1%
$150,000 to $199,999	12.0%	8.7%	10.52	0.0012	8.5%	15.19	<0.0001	6.8%
$200,000 or more	12.0%	11.6%	0.12	0.73	10.3%	3.17	0.075	7.7%

n represents the sample size
df=1
**The highlighted rows indicate significant differences*

No differences existed between the average visitor demographics of nine POIs and the 2018 Visitor Use Surveys regarding gender. Only two age subgroups (i.e., 25-year-old to 34-year-old and 75-year-old and older) showed a statistically significant difference between the survey and SafeGraph data. There was also a statistically significant difference in Asians in the race category. In addition, education level and income level showed statistically significant differences between the two data sources, and only two income subgroups ($50,000 to $74,999 and $200,000 or more) lacked significant differences.

Table 4 presents the top ten states of visitor origins from SafeGraph data and the survey results. Slight differences existed between the ranks of states. For example, Wyoming (WY) is ranked in the top 10 only in the SafeGraph data, while New York (NY) is ranked in the top 10 only in the survey. Additionally, the ranks of states by the two datasets had different

Table 4
Comparison of origins of states of visitors between 2018 SafeGraph data (combining with ACS) and Yellowstone National Park Summer 2018 Visitor Use Surveys.

2018 Visitor Use Surveys		2018 SafeGraph Nine POIs' Average		2018 Yellowstone Arts & Photography Center	
State	Percentage	State	Percentage	State	Percentage
CA	13.0%	CA	9.0%	CA	10.8%
TX	5.0%	WY	6.8%	ID	6.4%
FL	4.0%	TX	5.8%	UT	5.2%
WA	4.0%	MT	5.1%	TX	5.2%
UT	4.0%	FL	4.3%	MT	4.7%
CO	4.0%	UT	4.0%	WY	3.8%
NY	4.0%	ID	3.5%	CO	3.4%
MN	3.0%	WA	3.2%	FL	3.1%
ID	3.0%	MN	3.1%	OH	3.1%
MT	3.0%	OH	3.1%	WA	3.0%

orders.

4.2. Visitor temporal patterns

4.2.1. Monthly visits

Two POIs in SafeGraph data, Yellowstone Art & Photography Center and Old Faithful General Store, were selected to compare monthly visits with official statistics, since the two were the most visited POIs from 2018 to 2020. These two POIs have close proximity to Old Faithful Geyser, and both POIs showed similar peak periods (Fig. 2) (see Fig. 3).

The Pearson's r correlation coefficient of SafeGraph data of Yellowstone Art & Photography Center and recreation visits by NPS was 0.88 ($p < 0.001$), and the Pearson's r correlation coefficient of the Old Faithful General Store and official visitation statistics was 0.85 ($p < 0.001$) (Table 3), indicating that SafeGraph data and the NPS statistics have a strong correlation (The Odum Institute, 2015) regarding monthly visitation patterns.

4.2.2. Daily visit patterns

4.2.2.1. Trail counts from Old Faithful East vs. SafeGraph Data of Old Faithful General Store. The Old Faithful area was selected to validate daily visitation patterns of SafeGraph data. YNP staff installed a trail counter to collect visitor counts at the trailhead of Old Faithful East. Old Faithful General Store is the closest POI to the trailhead. The straight-line distance between the two points is approximately 700 m. In addition, the Old Faithful area has good cellular service to track mobile devices and capture visitation numbers.

Fig. 4 presents the daily visitation patterns and correlations of the two data sources. The Pearson's r correlation coefficients were 0.76 ($p < 0.001$) and 0.77 ($p < 0.001$) in 2018 and 2019 (Fig. 4), indicating strong relationships between the two data sources of daily visitation patterns.

4.2.2.2. Parkwide gate vehicle counts vs. SafeGraph Data of Yellowstone Arts & Photography Center. The POI, Yellowstone Arts & Photograph Center, the most visited POI, was selected to validate daily visits with Parkwide Gate Vehicle Counts. Fig. 5 presents the visitation correlations of the two data sources, showing a noticeable mismatched pattern from August 2020 to October 2020. The Pearson's r correlation coefficients were 0.79 ($p < 0.001$) and 0.47 ($p < 0.001$) in 2019 and 2020, respectively (Fig. 5).

In addition, since Yellowstone Arts & Photograph Center only opens during the summer season, the two data sources in 2019 were split into winter season (05/18/2019 to 06/20/2019 and January 10, 2019 to 10/17/2019) and summer (06/21/2019 to 09/30/2019, opening dates of the Center) seasons (Yellowstone Forever, 2019) and examined by Pearson's r correlation respectively. The result indicated that the coefficient of Pearson's r correlation between the two data sources for the winter season ($r = 0.87$) was stronger that the coefficient ($r = 0.70$) for the summer season (Fig. 6). This indicates that the summer opening season has active WiFi service, and more mobile devices were captured by SafeGraph (a steeper regression line) than those in the winter (a more gradual line).

5. Discussions

This study examined the use of mobile device data for studying national park visitor behavior and characteristics from three approaches: the accuracy of national park POIs, the estimated visitor demographics and origins, and the temporal national park visitation patterns. Only 40 POIs matched their location names and geographical locations among 80 POIs inside YNP as reported in SafeGraph data. The majority of those valid POIs are service facilities (e.g., Yellowstone Art & Photography Center, Old Faithful General Store, Canyon Lodge Cafeteria) that fall into YNP cellular coverage areas (Wadzinski, 2019), while attractions (e.g., Old Faithful Spring and Mammoth Hot Springs) or trails were invalid or not represented. Poor signal coverage in many areas of the park and a focus of SafeGraph on its business clientele, is a plausible reason to explain these results (Miyasaka et al., 2018; Muñoz et al., 2019).

Mixed results were found regarding visitor demographic comparisons. Mobile device data results revealed statistically significant differences with the results from the visitor use survey regarding education levels and income levels. Monz et al. (2021) obtained a similar result, suggesting that StreeLight Data were significantly different than traditional surveys regarding visitors' education levels. However, in our examination, no statistically significant differences existed between the two in terms of visitor gender distribution. Only two sub-groups of age (i.e., 25-year-old to 34-year-old, and 75-year-old and older) had

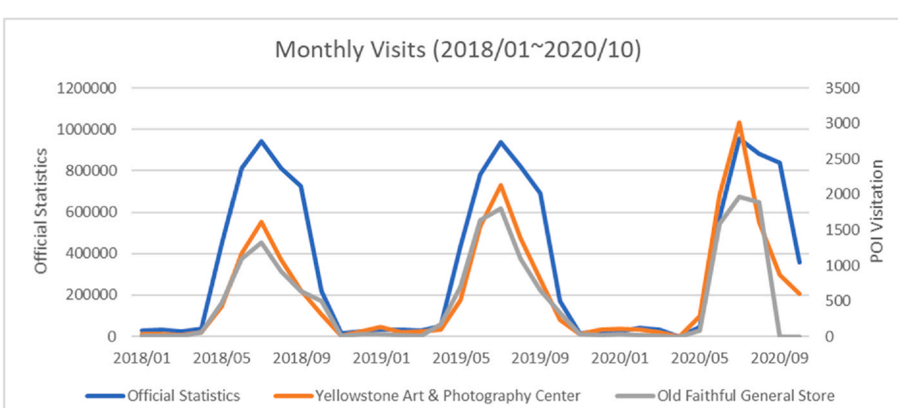

Fig. 2. Trends of monthly visits between NPS statistics and SafeGraph data (two POIs: Yellowstone arts & photography center and old faithful general store).

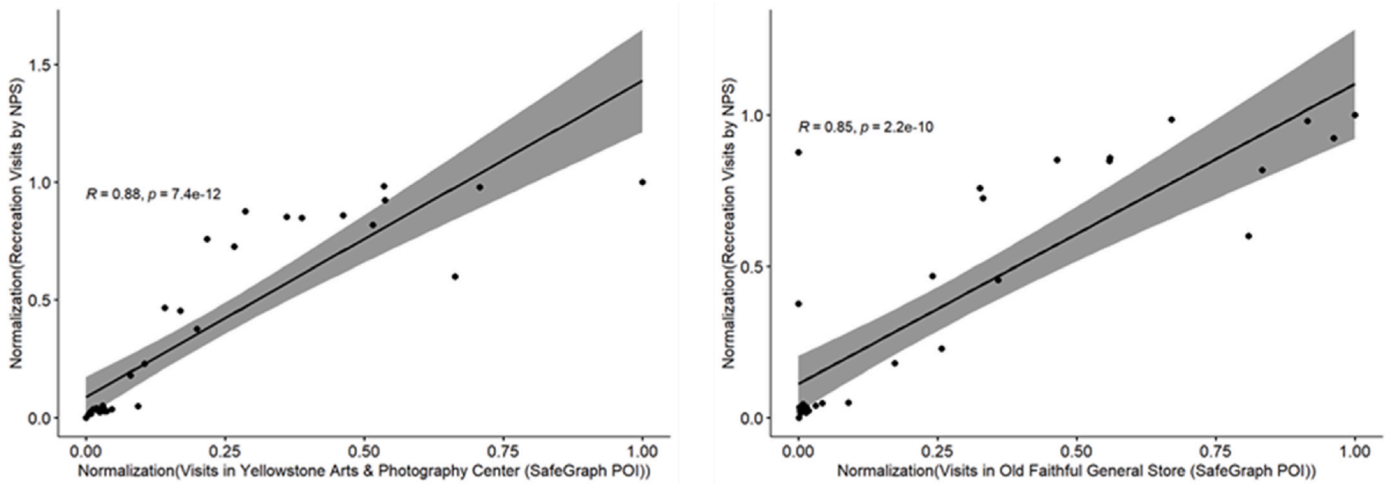

Fig. 3. Correlations of Monthly Visits between NPS Statistics and SafeGraph Data (Two POIs: Yellowstone Arts & Photography Center (Left) and Old Faithful General Store (Right)). *Note*. The shaded areas represent the 95% confidence interval.

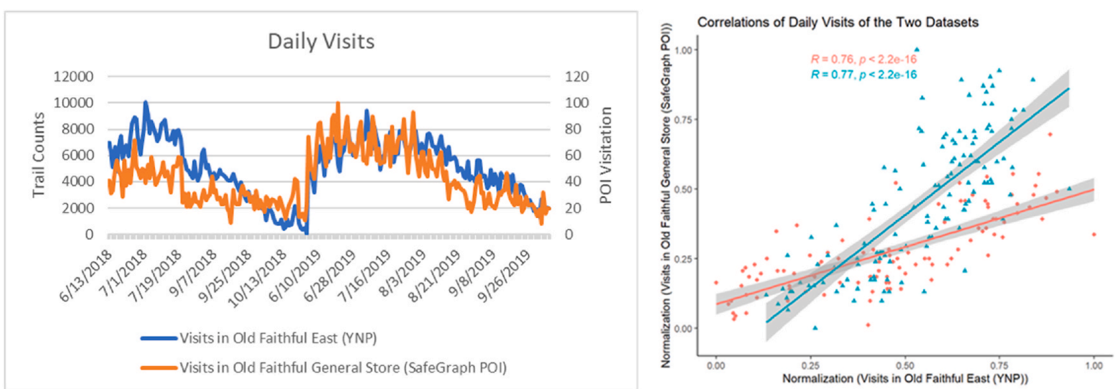

Fig. 4. Patterns (Left) and Correlations (Right) of Daily Visits between Trail Counters in Old Faithful East and SafeGraph POI in Old Faithful General Store. *Note*. The shaded areas (right) represent the 95% confidence interval.

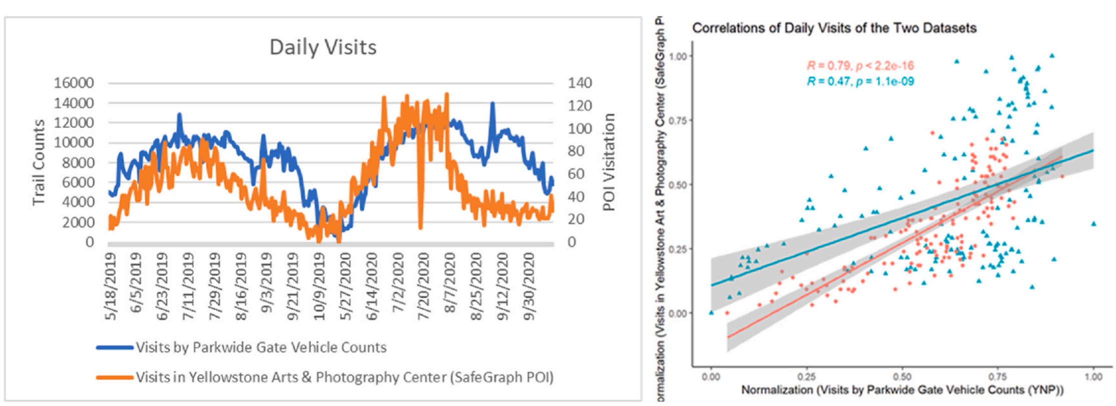

Fig. 5. Patterns (Left) and Correlations (Right) of Daily Visits between Parkwide Gate Vehicle Counts and Yellowstone Arts & Photography Center *Note*. The shaded areas (right) represent the 95% confidence interval.

statistically significant differences between the mobile device data and the survey.

There are several potential explanations for the differences in visitor demographics discovered in this research. SafeGraph data only covered domestic visitors and long-term international visitors, as each visitor is assigned a home CBG, while the traditional onsite park survey collected data from domestic, short-term, and long-term international visitors. According to the Visitor Use Surveys (National Park Service, 2019), 77% of visitors were from North America, 13% of visitors were from Europe, such as Germany, France, Switzerland, and the U.K, followed by visitors 8% from Asia (89% of all Asian visitors were Chinese). The survey reported that 17.0% of visitors were Asian or Asian Americans in the survey; in comparison, only 5.2% of visitors were Asian American based on SafeGraph data. This fact may have contributed to the statistically

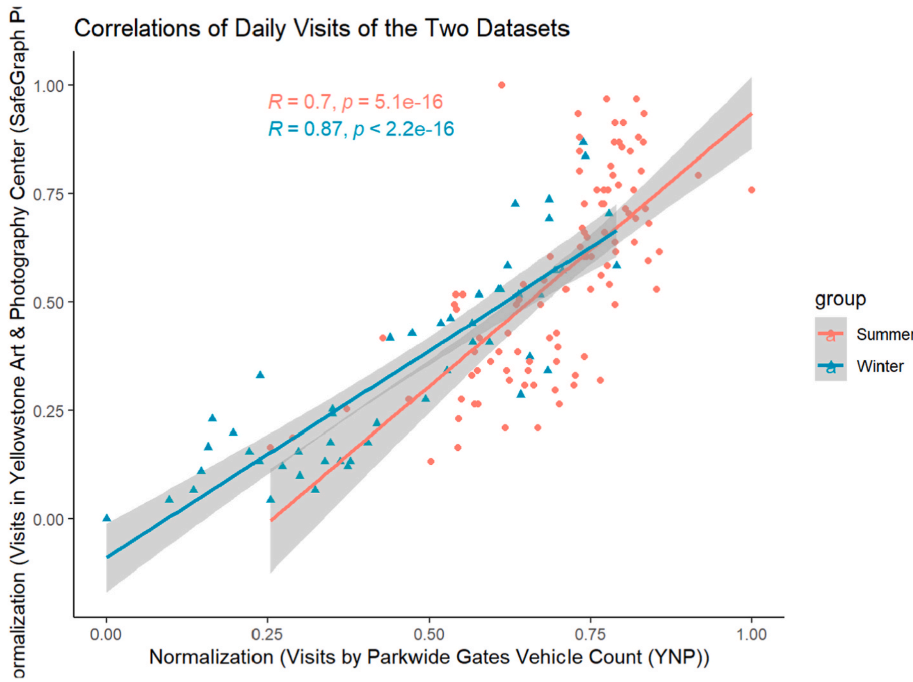

Fig. 6. Correlations of Daily Visits between Parkwide Gate Vehicle Counts and Yellowstone Arts & Photography Center for 2019 winter and 2019 summer *Note*. The shaded areas represent the 95% confidence interval.

significant differences in Asian percentages. These differences also highlight possible limitations of only using SafeGraph data to answer race-related questions.

Slight differences in the ranks of visitor origins existed between the two datasets. A possible reason why WY is ranked among one of the top 10 states of visitor origin based on SafeGraph data is that the mobile data capture visitors regardless of the nature of their visits. Therefore, locals, or even YNP managers/staff who regularly come to the park, could be counted as visitors. It is legitimate to count locals that may frequent the park as visitors, but the inclusion of YNP managers/staff is a limitation of SafeGraph data.

The mobile device data also show strong correlations with official count data in terms of temporal visitation patterns, except for the 2020 summer, during the COVID-19 pandemic. A potential reason for the low correlation was that Yellowstone Arts & Photography Center implemented indoor capacity limits due to the COVID-19 pandemic,[4] while Parkwide Gate Vehicle Counts estimated the visitation of all types of vehicles for the entire national park. Taff et al. (2022) reported similar regulations in that capacity limits were enforced in some facilities within five U.S. national parks to keep visitors and staff safe during the pandemic. Therefore, the visitation pattern of one specific attraction may not match the visitation patterns of the entire park and could be influenced by internal factors, such as pandemic-specific policies and associated visitor behaviors (Taff et al., 2022).

Additionally, the differences between the mobile device data and the count data can be explained by three potential reasons. First, Parkwide Gate Vehicle Counts counted all types of vehicles, meaning that the dataset contained counts of the employees, non-recreation, and recreational vehicles. In addition, the number of passengers in each vehicle varied. Second, although Yellowstone Arts & Photography Center was the most popular attraction among all the valid POIs, not all visitors visited this attraction. Additionally, the center only opens during the summer season. Third, SafeGraph partners, various mobile applications, obtained opt-in consent from its users to collect anonymous location data. Some users may have denied the tracking request from the mobile applications. Therefore, both parkwide gate vehicle counts and SafeGraph data may represent different visitor populations and were different from the actual number of visitors at the park level.

5.1. Implications

The results of the validation of mobile device data from the three approaches imply that adopting mobile device data, especially POIs and visitor demographics, should be approached with caution due to the limitations and potential external factors impacting the validity of these datasets. As reported in this study, half of the selected POIs inside YNP were invalid, and the majority of the invalid POIs were attractions. The results suggest that researchers or park managers should carefully select POIs with accurate geographical locations before assessing visitor demographics and temporal visitation patterns. The open dates and specific policies of attractions and facilities should also be considered when selecting appropriate POIs.

For estimating visitor demographics, SafeGraph assigns a home location for a mobile device and determines visitors' home CBGs by analyzing six weeks of data during nighttime hours. Although international visitor data are trackable by SafeGraph data, which are aggregated into raw counts of visitors in the *Patterns* dataset, this approach cannot provide home CBGs for short-term international visitors. According to the Visitor Use Surveys, international visitors accounted for 23% of all visitors to YLP (National Park Service, 2019). The proportion of international visitors will impact the accuracy of assessing visitor demographics with SafeGraph data. Therefore, it would be better to apply mobile device data to estimate visitor demographics in national parks with more domestic visitors (e.g., Gettysburg National Military Park; Cuyahoga Valley National Park, etc.). Further validation studies are needed for other national parks.

When researchers utilize mobile device data to estimate temporal visitation patterns, it is necessary to consider other factors that may influence visitor behaviors. For example, during the beginning of the COVID-19 pandemic, some service facilities limited the number of visitors, which altered visitor flows and caused significant differences in

[4] This information was confirmed by a social scientist from YNP.

visitation numbers compared to regular periods (Taff et al., 2022).

Although the quality of mobile device data still needs to be improved and approached with consideration of the limitations – especially in the geographical accuracy of POIs and visitor demographics – national parks lacking enough staff and financial support can benefit from this type of data to roughly estimate visitor characteristics and temporal visitation patterns. We see the use of mobile device data as a significant contribution to management of many types of parks and protected areas.

5.2. Limitations

Four limitations potentially impact the results of this study. First, SafeGraph data contain platform bias since it only captures information from mobile application users who grant those applications to collect their anonymous location data. These users only compose a fraction of mobile phone users. The populations tracked by SafeGraph and those visiting YNP could be different. The results demonstrated that they are not a random sample of all Yellowstone visitors. Therefore, changes in the opt-in process of mobile apps could impact data quality. For example, starting in April 2021, when iPhone users open any apps that want to access their device I.D, they will be asked if they want to be tracked and are given an opportunity to opt out (Leswing, 2021). This new privacy setting could increase the opt-out ratios and change the demographics represented in mobile phone data.

Secondly, in this study, the majority of selected POIs were service facilities, not attractions of YNP; therefore, one bias of the represented POIs is that the characteristics of those who visited service facilities could be different from those who visited attractions or trails in YNP. Furthermore, different selected POIs could attract visitors with different characteristics.

Thirdly, YNP have different operating hours and seasons for various attractions or services within the park (Park, 2022b), which has impacts on estimating temporal visitations. For example, the visitation in Yellowstone Arts & Photography Center by SafeGraph has a stronger relationship with gate vehicle count in the winter season than in the summer season: the availabity of WiFi service may have complicated the accuracy of SafeGraph data in the summer. In addition, using POIs to assess visitor behaviors is also challenging during the COVID-19 pandemic or future large-scale events that may significantly alter visitor behaviors. For instance, Fig. 5 shows a low correlation between SafeGraph data and gate vehicle counts in 2020 since Yellowstone Arts & Photography Center implemented indoor capacity limits for visitors. Additionally, the mask requirement policy in all NPS buildings could also reduce visitation volumes (Park, 2021b). Therefore, internal factors, such as seasons or hours of operation and pandemic-specific policies of an attraction, could influence the estimated visitation patterns based on mobile device data (Taff et al., 2022).

Fourthly, the target population of the Visitor Use Surveys included both domestic and international visitors, whereas SafeGraph data only captured domestic and long-term international visitors. SafeGraph data would have likely been more representative during the 2020 YNP visitation cycle, as short-term international visitation decreased significantly during that pandemic-impacted years. Still, these data would not have been as representative during more normal years.

Therefore, based on the discussion above, additional information, such as visitor characteristics and policies for park facilities, collected by surveys or other types of approaches are indispensable to supplement mobile device data.

5.3. Future research

This study highlights several areas to further advance research regarding mobile device data. First, researchers should collaborate with the companies that provide mobile device data to improve data quality, as this study revealed that 40 POIs are invalid after the POI validation process. Using POIs with inaccurate geolocations could mislead park managers regarding the spatial distribution of visitors. Improved data quality can benefit both national park researchers and these commercial companies.

Second, future research could employ visitor demographics beyond the five demographic variables and origins of visitor residency and investigate how these variables influence temporal visitation patterns (Rice and Pan, 2020). For example, ACS includes many other variables, such as unemployment rates. In addition, although this study calculated the average visitor demographics of the nine POIs and the most visited POI, it is feasible to retrieve the visitor demographic for each POI. This suggests that future research could investigate and compare visitor demographics for various attractions and service facilities and understand different segments' unique preferences.

Third, park managers can potentially utilize mobile device data to track changes in visitation numbers and demographics at a park level or at a specific attraction over time and gain insights for adaptive management strategies (Monz et al., 2019, 2021). Using experimental designs, such as implementing different staffing or infrastructural adjustments as various conditions, researchers could examine behavioral changes using mobile device data. Understanding mobile device data and adopting them could help park managers understand the effects of various adaptive management strategies.

Furthermore, future research can combine mobile device data with other data sources to understand visitor behaviors beyond ACS. For example, the linkage of mobile device data with textual content from social media data can help researchers and park managers to understand visitor motivations, constraints, and positive/negative experiences in national parks. In addition, linking mobile device data with point-of-sale data can reveal the economic impacts of visitors to various service facilities in national parks (Marques et al., 2022) and in gateway communities. Moreover, mobile device data can not only be utilized for visitor use in remote national parks but also to explore visitor behaviors and characteristics in urban parks and community parks. The utility of mobile phone data might be higher for the latter due to better cell coverage in more populated areas.

Finally, investigating the impact of the COVID-19 pandemic on national park visitation with mobile phone data could be a promising direction, given that most onsite visitor survey projects were not possible when quarantine and social distancing policies were in place (Taff et al., 2022).

6. Conclusions

This study assessed the validity of mobile device data for validating POIs, visitor demographics, and temporal visitation patterns in YNP by comparing it with traditional approaches, including onsite visitor use and count data. The similarities and differences between the two data sources suggest that SafeGraph data can serve as an additional and complementary source of information to traditional methods with careful consideration. With the advancement of technology in the future and more in-depth validation, mobile device data could provide more detailed, comprehensive, and timely information related to visitor characteristics and temporal visitation patterns for national parks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the National Science Foundation (Award # SES-1823633), the USDA National Institute of Food and Agriculture and Multistate Research Project #PEN04623 (Accession #1013257), and the Social Science Research Institute and the Institute of Computational and Data Sciences of Penn State University. The views expressed in this paper are the authors' responsibility and do not necessarily represent the opinions or policy of the National Park Service.

References

Alba, C., Pan, B., Yin, J., Rice, W.L., Mitra, P., Lin, M., Liang, Y., 2022. COVID-19's Impact on Visitation Behavior to U.S. National Parks from Communities of Color—Evidence from Mobile Phone Data. https://doi.org/10.21203/rs.3.rs-1182484/v1.

Bonack, B., 2021. SafeGraph's Data Sourcing Process. March 26. https://www.safegraph.com/blog/safegraphs-data-sourcing-process.

Cessford, G., Muhar, A., 2003. Monitoring options for visitor numbers in national parks and natural areas. Journal of Natural Conservation 11, 240–250.

Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 589 (7840), 82–87. https://doi.org/10.1038/s41586-020-2923-3.

Chiou, L., Tucker, C., 2020. Social Distancing, Internet Access and Inequality (No. W26982). *National Bureau of Economic Research*. https://doi.org/10.3386/w26982.

Creany, N.E., Monz, C.A., D'Antonio, A., Sisneros-Kidd, A., Wilkins, E.J., Nesbitt, J., Mitrovich, M., 2021. Estimating trail use and visitor spatial distribution using mobile device data: an example from the nature reserve of orange county, California USA. Environmental Challenges 4, 100171. https://doi.org/10.1016/j.envc.2021.100171.

Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front. Environ. Sci. 3 https://doi.org/10.3389/fenvs.2015.00063.

English, D., Bowker, J.M., 2018. Introduction to the special issue on visitor monitoring. J. Park Recreat. Adm. 36, IX–X.

Forever, Yellowstone, 2019. New Art Exhinits In Yellowstone. June 21. https://www.yellowstone.org/new-art-exhibits-in-yellowstone/#:~:text=The%20Yellowstone%20Art%20%26%20Photography%20Center%20is%20located%20between%20the%20Visitor,a.m.%20to%206%3A00%20p.m.

Gstaettner, A.M., Lee, D., Weiler, B., 2020. Responsibility and preparedness for risk in national parks: results of a visitor survey. Tour. Recreat. Res. 45 (4), 485–499. https://doi.org/10.1080/02508281.2020.1745474.

Hadwen, W.L., Hill, W., Pickering, C.M., 2007. Icons under threat: why monitoring visitors and their ecological impacts in protected areas matters. Ecol. Manag. Restor. 8 (3), 177–181. https://doi.org/10.1111/j.1442-8903.2007.00364.x.

Juhasz, L., Hochmair, H., 2020. Studying Spatial and Temporal Visitation Patterns of Points of Interest Using SafeGraph Data in Florida. https://doi.org/10.1553/giscience2020_01_s119. GIS Center.

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., Kruse, J., 2020. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Sci. Data 7 (1), 390. https://doi.org/10.1038/s41597-020-00734-5.

Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., Shirayama, Y., 2020. Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. Tourism Manag. 77, 104010. https://doi.org/10.1016/j.tourman.2019.104010.

Kupfer, J.A., Li, Z., Ning, H., Huang, X., 2021. Using mobile device data to track the effects of the COVID-19 pandemic on spatiotemporal patterns of national park visitation. Sustainability 13 (16), 9366. https://doi.org/10.3390/su13169366.

Laney, D., 2001. 3D data management: controlling data volume, velocity and variety. META Group Research Note 6, 70.

Lawson, S., Chamberlin, R., Choi, J., Swanson, B., Kiser, B., Newman, P., Gamble, L., 2011. Modeling the effects of shuttle service on transportation system performance and quality of visitor experience in Rocky Mountain National Park. Transport. Res. Rec. 2244 (1), 97–106.

Lawson, S.R., Newman, P., Monz, C., 2017. A systems-based approach to address unintended consequences of demand-driven transportation planning in national parks and public lands. International Journal of Sustainable Transportation 11 (2), 98–108.

Lee, M., Zhao, J., Sun, Q., Pan, Y., Zhou, W., Xiong, C., Zhang, L., 2020. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. PLoS One 15 (11), e0241468. https://doi.org/10.1371/journal.pone.0241468.

Leggett, C., Horsch, E., Smith, C., Unsworth, R., 2017. Estimating Recreational Visitation to Federally-Managed Lands (Cambridge, MA).

Leswing, K., 2021. Apple Will Start Enforcing the iPhone Privacy Change that Facebook Is Worried about Next Week. CNBC. April 20. https://www.cnbc.com/2021/04/20/apple-ios-14point5-release-date-with-att-idfa-restrictions-confirmed-for-next-week.html.

Lu, S., Fang, Z., Zhang, X., Shaw, S.-L., Yin, L., Zhao, Z., Yang, X., 2017. Understanding the representativeness of mobile phone location data in characterizing human mobility indicators. ISPRS Int. J. Geo-Inf. 6 (1), 7. https://doi.org/10.3390/ijgi6010007.

Ma, S., Kirilenko, A., 2021. How reliable is social media data? Validation of TripAdvisor tourism visitations using independent data sources. In: Wörndl, W., Koo, C., Stienmetz, J.L. (Eds.), Information and Communication Technologies in Tourism 2021. Springer International Publishing, pp. 286–293. https://doi.org/10.1007/978-3-030-65785-7_26.

Marques, C.P., Guedes, A.S., Bento, R., 2022. Tracking changes in tourism demand with point-of-sale data: the case of Portugal. Tourism Hospit. Res. 0 (0), 1–7. https://doi.org/10.1177/14673584221075175. In press.

Merrill, N.H., Atkinson, S.F., Mulvaney, K.K., Mazzotta, M.J., Bousquin, J., 2020. Using data derived from cellular phone locations to estimate visitation to natural areas: an application to water recreation in New England, USA. PLoS One 15 (4), e0231863. https://doi.org/10.1371/journal.pone.0231863.

Miyasaka, T., Oba, A., Akasaka, M., Tsuchiya, T., 2018. Sampling limitations in using tourists' mobile phones for GPS-based visitor monitoring. J. Leisure Res. 49 (3–5), 298–310. https://doi.org/10.1080/00222216.2018.1542526.

Monz, C., Mitrovich, M., D'Antonio, A., Sisneros-Kidd, A., 2019. Using mobile device data to estimate visitation in parks and protected areas: an example from the nature reserve of Orange County, California. J. Park Recreat. Adm. 37 (4) https://doi.org/10.18666/JPRA-2019-9899. Article 4.

Monz, C., Creany, N., Nesbitt, J., Mitrovich, M., 2021. Mobile device data analysis to determine the demographics of park visitors. J. Park Recreat. Adm. 39 (1) https://doi.org/10.18666/10.18666/JPRA-2020-10541.

Muñoz, L., Hausner, V.H., Monz, C.A., 2019. Advantages and limitations of using mobile apps for protected area monitoring and management. Soc. Nat. Resour. 32 (4), 473–488. https://doi.org/10.1080/08941920.2018.1544680.

National Park Service Stats, 2021. Recreation Visits by Month. February 27. https://irma.nps.gov/STATS/SSRSReports/Park%20Specific%20Reports/Recreation%20Visitors%20By%20Month%20(1979%20-%20Last%20Calendar%20Year)?Park=YELL.

Pappalardo, L., Pedreschi, D., Smoreda, Z., Giannotti, F., 2015. Using Big Data to Study the Link between Human Mobility and Socioeconomic Development. 2015 IEEE International Conference on Big Data (Big Data). Santa Clara, CA, USA, pp. 871–878. https://doi.org/10.1109/BigData.2015.7363835, 2015.

National Park Service Office of Communications, 2020a. National Park Visitation Tops 327 Million in 2019. February 27. https://www.nps.gov/orgs/1207/2019-visitation-numbers.htm#:~:text=WASHINGTON%20%E2%80%93%20America's%20national%20parks%20continue,record%20keeping%20began%20in%201904.

National Park Service, 2019. Yellowstone National Park Summer 2018 Visitor Use Surveys. November 07. https://www.nps.gov/yell/learn/management/upload/2018-Yellowstone-Visitor-Use-Surveys-FINAL-REPORT_WEB-RESOLUTION.pdf.

National Park Service, 2021b. National Park Service Implements National Mask Requirement. August 16. https://www.nps.gov/orgs/1207/covid-mask-update-aug-2021.htm.

National Park Service, 2022a. Visitation Numbers. February 16. https://www.nps.gov/aboutus/visitation-numbers.htm.

National Park Service, 2022b. Operating Hours & Seasons. May 17. https://www.nps.gov/yell/planyourvisit/hours.htm.

Park, S., Xu, Y., Jiang, L., Chen, Z., Huang, S., 2020. Spatial structures of tourism destinations: a trajectory data mining approach leveraging mobile big data. Ann. Tourism Res. 84, 102973. https://doi.org/10.1016/j.annals.2020.102973.

Pettebone, D., Meldrum, B., 2018. The need for a comprehensive socioeconomic research program for the National Park Service. George Wright Forum 35 (1), 22–31.

Pettebone, D., Newman, P., Lawson, S.R., 2010. Estimating visitor use at attraction sites and trailheads in Yosemite National Park using automated visitor counters. Landsc. Urban Plann. 97 (4), 229–238. https://doi.org/10.1016/j.landurbplan.2010.06.006.

Pettebone, D., Meldrum, B., Leslie, C., Lawson, S.R., Newman, P., Reigner, N., Gibson, A., 2013. A visitor use monitoring approach on the Half Dome cables to reduce crowding and inform park planning decisions in Yosemite National Park. Landsc. Urban Plann. 118, 1–9. https://doi.org/10.1016/j.landurbplan.2013.05.001.

Pettebone, D., D'Antonio, A., Sisneros-Kidd, A., Monz, C., 2019. Modeling visitor use on high elevation mountain trails: an example from Longs Peak in Rocky Mountain National Park, USA. J. Mt. Sci. 16 (12), 2882–2893. https://doi.org/10.1007/s11629-019-5663-9.

Pew Research Center, 2021. Demographics of Mobile Device Ownership and Adoption in the United States. April 7. Retrieved March 23, 2022, from. https://www.pewresearch.org/internet/fact-sheet/mobile/.

Phithakkitnukoon, S., Smoreda, Z., Olivier, P., 2012. Socio-Geography of human mobility: a study using longitudinal mobile phone data. PLoS One 7 (6), e39253. https://doi.org/10.1371/journal.pone.0039253.

Raun, J., Ahas, R., Tiru, M., 2016. Measuring tourism destinations using mobile tracking data. Tourism Manag. 57, 202–212. https://doi.org/10.1016/j.tourman.2016.06.006.

Rice, W.L., Pan, B., 2020. Understanding drivers of change in park visitation during the COVID-19 pandemic: A spatial application of big data. https://doi.org/10.31235/osf.io/97qa4. *SocArXiv*.

Rice, W.L., Park, S.Y., Pan, B., Newman, P., 2019. Forecasting campground demand in U. S. national parks. Ann. Tourism Res. 75, 424–438. https://doi.org/10.1016/j.annals.2019.01.013.

Rice, W.L., Rushing, J., Thomsen, J., Whitney, P., 2022. Exclusionary effects of campsite allocation through reservations in U.S. National parks: evidence from mobile device location data. J. Park Recreat. Adm. https://doi.org/10.18666/JPRA-2022-11392.

Rodríguez, J., Semanjski, I., Gautama, S., Van de Weghe, N., Ochoa, D., 2018. Unsupervised hierarchical clustering approach for tourism market segmentation based on crowdsourced mobile phone data. Sensors 18 (9), 2972. https://doi.org/10.3390/s18092972.

SafeGraph, 2020. Places Manual. December 04. https://docs.safegraph.com/docs/places-manual.

SafeGraph, 2021. FAQs. May 14. https://docs.safegraph.com/docs/faqs.

Sessions, C., Wood, S.A., Rabotyagov, S., Fisher, D.M., 2016. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. J. Environ. Manag. 183, 703–711. https://doi.org/10.1016/j.jenvman.2016.09.018.

Taff, B.D., Thomsen, J., Rice, W.L., Miller, Z., Newton, J., Miller, L., Gibson, A., Riddle, M., Schaberl, J.P., McCormick, M., 2022. U.S. National park visitor experiences during COVID-19: data from acadia, glacier, grand teton, shenandoah, and Yellowstone national parks. Parks Stewardship Forum 38 (1). https://doi.org/10.5070/P538156128.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: assessing the usability of social media data for visitor monitoring in protected areas. Sci. Rep. 7 (1), 17615. https://doi.org/10.1038/s41598-017-18007-4.

The Odum Institute, 2015. Learn about pearson's correlation coefficient in SPSS with data from the global health observatory data (2012). In: SAGE Research Methods Datasets Part 1. SAGE Publications, Ltd. https://doi.org/10.4135/9781473948167.

Traag, V.A., Browet, A., Calabrese, F., Morlot, F., 2011. Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference. *2011 IEEE Third International Conference On Privacy, Security, Risk And Trust And 2011*. IEEE Third International Conference on Social Computing, Boston, MA, USA, pp. 625–628. https://doi.org/10.1109/PASSAT/SocialCom.2011.133, 2011.

United States Census Bureau. (n.d.). North American Industry Classification System. https://www.census.gov/naics/.

United States Census Bureau, 2020a. September 3 Understanding Geographic Identifiers (GEOIDs). https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html.

United States Census Bureau, 2020b. American Community Survey (ACS) 2015-2019 (5-Year Estimates). December 10. https://www.socialexplorer.com/data/ACS2019_5yr.

Wadzinski, G., 2019. *Do Cell Phones Work In Yellowstone? Is There WiFi?* Yellowstone National Park Trips. December 2. https://www.yellowstonepark.com/park/faqs/cell-phone-wifi-yellowstone/.

Ziesler, P.S., Pettebone, D., 2018. Counting on visitors: a review of methods and applications for the National Park Service's visitor use statistics program. J. Park Recreat. Adm. 36 (1) https://doi.org/10.18666/JPRA-2018-V36-I1-8104.