# Evaluating Mobility Data for Recreation Monitoring: Diagnostics and Implications for Best Practice

Cameron Duff – National Oceanic and Atmospheric Administration, Assessment and Restoration Division, Silver Spring, MD 20910, USA

Alice Antia – Industrial Economics, Incorporated, Cambridge MA 02140, USA

James Hilger – National Oceanic and Atmospheric Administration, Assessment and Restoration Division, Silver Spring, MD 20910, USA

Eric Horsch – Industrial Economics, Incorporated, Cambridge MA 02140, USA

Nathaniel Merrill – Matunuck Research Group, South Kingstown RI, USA

Jason Murray – National Oceanic and Atmospheric Administration, Assessment and Restoration Division, Silver Spring, MD 20910, USA

Sama Winder – Outdoor R&D Lab, University of Washington, Seattle, WA 98195, USA

Spencer Wood – Outdoor R&D Lab, University of Washington, Seattle, WA 98195, USA

Matthew Zafonte – California Department of Fish and Wildlife, Office of Spill Prevention and Response



Abstract

Mobile device location data (or mobility data, MD) are a novel and exciting source of information for recreational monitoring. In this paper we analyze datasets from four commercial vendors across three case studies to illustrate some important challenges. Our results show inconsistent spatial-temporal patterns and correlations with on-site counts. Given this evidence, we describe a conceptual model of the potential sources of error and structural variability in visitation estimates derived from MD. We discuss approaches for evaluation of data quality and suggest a range of supplemental data products that vendors could provide to support recreation analysis.

NOAA Disclaimer: *The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect the views of NOAA or the Department of Commerce.*

doi:10.3368/le.102.3.011626-0006

The authors wish to thank Jeff Beauvais and participants in the workshop sponsored by the editors of *Land Economics*, the editors themselves and two anonymous referees for thoughtful comments and suggestions.

# 1. Introduction

Commercially available mobility data (MD)[i] present an opportunity for researchers interested in understanding outdoor recreational behavior. In recent years, the proliferation of smartphone users and applications has generated vast amounts of MD, and in turn an emerging industry of vendors who aggregate and sell these products for research and commercial applications. There are over a billion acres of federal lands and waters alone designated for recreation uses (Leggett et al. 2017), not to mention state, local, and other properties, and many resource managers require estimates of recreational visitation to effectively manage the lands and waters, as well as visitor experiences. For this reason, MD provides an opportunity to measure and characterize recreational visitation, potentially replacing or enhancing traditional data collection methods (Tsai et al. 2023, Merrill et al. 2025).

Researchers have explored this opportunity in a variety of public land and waters applications, resulting in a growing body of work that relies on MD as a proxy for visitation (e.g., Csomos et al. 2023), to explore visitor demographics (e.g., Zhang et al. 2024), estimate recreation value (e.g., Gonyo et al. 2024), and reveal visitor preferences (e.g., Song et al. 2022), among other applications. Indeed, many of the other papers in this special issue use MD to address a variety of questions related to visitor estimation, demand estimation equations, zero market-share adjustments, etc. However, MD are rarely validated against traditional measures of visitation that are understood to be reliable (Hanson et al. 2025), and research which does conduct these evaluations has found that the relationships between these traditional data and MD vary across time, space, and mobile provider (Monz et al. 2018, Filazzola et al. 2022, Liang et al. 2022, Tsai et al. 2023, Parkinson et al. 2025, Winder et al. 2025). Further, MD are rarely validated using within-dataset comparisons over time and space, which this paper presents.

Multiple causes likely explain the poor track record of MD in these evaluations. To start, data quality can vary considerably across the different MD vendors active in the market (Winder et al. 2025). The data generating process and complex processing decisions made by MD vendors are considered trade secrets. This lack of transparency means a researcher cannot replicate a mobility dataset or truly understand the quality of the product they have obtained. Further, these datasets have evolved considerably over time due to changes in the proprietary methods, data vendor companies and their business models, consumer protection laws, and smartphone user behavior, as well as other factors (e.g., the extent of fraudulent data). Our experience using these data suggests these changes do not demonstrate consistent improvement in data quality.

All sampling schemes have some degree of measurement error, and in the best-case scenario we expect all the typical complications (see Heckman 1979) in our statistical inferences based on MD. This is also generally true for the reference data we use for comparison. Specific components of our conceptual model (see Table 1) have connections to classical (e.g., spatial imprecision of estimates) and non-classical (e.g., willingness to share location data is correlated with demographics) measurement error, but overall, we focus on ideas related to sampling error specifically, not measurement error broadly. Sampling error and issues of representativeness are a direct consequence of the numerous and opaque upstream processing steps in the delivery of a MD product. Consider the example of heterogeneity in willingness to share location data. None of the vendors we have spoken with offer individual-level demographic data, even though it is likely an important determinant of the sampling rate. For instance, Struminskaya et al. (2021) highlight the importance of demographics in willingness to share GPS location data with researchers. This challenge also has a clear connection to non-response bias in the survey literature. Dutwin et al. (2017) show the non-response rate varies not only by demographics, but also that this relationship varies across technologies - a relationship that is underexplored in the MD literature. In the discussion section, we highlight some ways in which MD vendors can help researchers to better account for sampling error.

This paper focuses on the topic of critically evaluating mobility data for the purpose of recreation monitoring. While our case studies and discussion are primarily oriented toward the task of accurately calculating visitation, the conclusions extend directly to recreational demand models. For example, key parameters in recreational demand analysis (price elasticities, substitution patterns, travel cost) depend on accurately measuring the underlying behavior of recreators. Our paper highlights how the use of MD introduces additional sources of classical and non-classical measurement error in the calculation of these parameters.

Using case studies with datasets from four MD vendors, we demonstrate a number of approaches for interrogating the quality of a MD source[ii]. We also propose common language for these problems so that researchers may communicate more effectively moving forward. Given the evaluation of MD from our case studies, the paper describes a conceptual model of the potential sources of error and structural variability in a visitation model derived from MD. We argue these characteristics of MD are likely to be a common problem across MD vendors and merit consideration for virtually all recreation monitoring applications.

In Section 2, we describe the mobility datasets applied in this paper, along with reference datasets (traditional on-the-ground visitation data considered to be reliable measures of recreation use) used for comparison in some of our case studies. We explain the various ways that each vendor prepares and delivers their datasets. In Section 3, we present case studies from three settings. In the first two case studies, we focus on MD suitability for specific research tasks. In the third, we use a generic setting to investigate the quality of mobility data more broadly. In Section 4, we provide a conceptual model illustrating some pathways by which MD-based measures of recreation can systematically depart from reality. In Section 5, we summarize our findings, discuss the relative suitability of MD for some common applications, and suggest how the MD industry can better adapt to the needs of the research community.

# 2. Datasets by Case Study

Datasets from four different mobility data vendors are applied to our three case studies, as summarized in Appendix Table A1. Each of the vendors provide data in a slightly different format[iii], but each can be reprocessed into unique daily visitors to a custom defined area of interest (AOI) (Wood et al. 2013), the common unit of analysis in this paper and in many MD applications.

Data from Provider 1 (used only in the first case study) are the most heavily processed of the four, offering AOI-level daily “count” data alongside an “estimate” count, modeled by the vendor to reflect true (population-level) visitation. Our case study features only the “estimate” variable for reasons described in the corresponding section. The Provider 2 data are more disaggregated, with individual pings combined into algorithmically defined “activities” from a single device. As a result, we often observe multiple activities from the same device in a site-day, which are further combined to form a count of unique daily visitors. One data evaluation advantage to this level of aggregation is that it allows researchers to plot activities on a map and examine the spatial distribution. Provider 3 has the same data format as Provider 2. The

Provider 4 dataset is the most disaggregated of the four, with the unit of observation being a single “ping” of location and timestamp recorded by a phone. This level of disaggregation can be helpful for additional customization in how activities or unique daily visitors are calculated, but it also requires additional geographical data processing steps.

Our analyses cannot make exact evaluative comparisons between the four vendors. These datasets were pulled together from different research groups and obtained at different points in time. The earlier datasets from Providers 1 and 2 are final products that cannot be re-queried over additional points in space, as the subscriptions have lapsed. Providers 3 and 4 offered a limited back-catalog that did not reach back far enough in time to make a full comparison. For the periods of time where we have overlap in areas of spatial analysis (e.g., Figures 2 and 6), the results do not suggest a recommendation of one over the other. This is a productive goal for future analysis, but it is outside the scope of this paper.

We utilize three reference data sources corresponding to the three case studies. In the first, we use Port Chilkoot cruise ship dock data from Haines, Alaska. The data were obtained from the 2023 Haines Borough Tourism Department Seasonal Report and describe yearly landings of ships, passengers and crew from the 2019 and 2022 seasons. In the second case study, we use daily counts of paid vehicles from 2019 to 2022 at several California State Park beaches where parking is closely related to total site recreation and managers use reliable and consistent counting methods. Finally, in the third case study, we use Major League Baseball (MLB) attendance data for 2019 to 2024 that were obtained from baseball-reference.com, representing the number of tickets sold for a given day (with occasional double-headers omitted from analysis for simplicity).



## 3. Case Studies

In this section, we present case study results using the mobility and reference data sources described in Section 2. Collectively, these datasets cover four prominent MD vendors and include data from 2019-2024 in different contexts. Through these case studies, we identify problematic data issues with MD and make relative statements about dataset reliability. We avoid making universal claims about whether a poor result in any given evaluation invalidates a dataset’s reliability because there may be research applications for which the data issues are acceptable and MD still suitable. However, these are heuristic starting points for such an evaluation. The reliability of a mobility dataset appears to vary considerably according to settings, time frame, and even when it was queried[iv]. Our case studies provide some potentially useful approaches for evaluating the quality of a mobility dataset in future applications.

The case studies compare MD with reliable reference datasets and make within-dataset comparisons of MD over time and space. Our first two case studies are motivated by specific research questions in Haines, AK and Southern California, as described in the sections that follow, and focus on the evaluation of datasets for those specific purposes. Our third case study presents a comparison of mobility datasets to MLB stadium attendance data simply for the purpose of evaluating MD rather than to answer any research questions about MLB attendance. A key question across these case studies is whether the patterns in the data are stable across years and exhibit expected responses to exogenous variables, namely weather and closed sites.



## 3.1 Port Chilkoot Dock - Data from Provider 1 and Alaska Cruises

As our first case study, we summarize research conducted by the Environmental Protection Agency (EPA) Office of Research and Development under a Recreation Economy for Rural Communities grant project in Haines, AK. Of interest to the town and project partners was understanding more about their visitors and how they used recreational areas as resource managers created plans to improve recreation-related economic opportunities. The time period of analysis was 2019 to 2022, so one important research question was how many visitors were returning to Haines after the COVID disruptions.

EPA procured a dataset from Provider 1 covering transportation hubs and recreational areas from 2019 to 2022. One important location included in the project was the Port Chilkoot cruise ship dock, which had reference data on landings. As Appendix Figure A1 shows, the site is situated with little other infrastructure or attractions nearby (red outline shows the geofence used for the MD query). Therefore, the Provider 1 data were expected to capture activity associated with the port and nothing else. This effort extended some earlier work that used a similar data product, which provided evidence of its usefulness in modeling visitation over a single summer period, based on its strong relationship with reference datasets (Merrill et al. 2020). The trends over a longer time period (multiple years) and at multiple sites were of interest to the Haines project funders.

In addition to the MD, EPA obtained cruise ship dock landing data from the 2023 Haines Borough Tourism Department Seasonal Report, describing yearly landings of ships, passengers, and crew from 2019 and 2022. To assess stability, EPA compared the 2019 and 2022 years from the MD to the dock landings data (see Appendix Table A2). EPA used Provider 1's proprietary "estimate" variable for this comparison, advertised as representing a population-extrapolated estimate for each day in the dataset, the desired metric for the project. It is generated using their proprietary algorithm to extrapolate from daily device "count" data, using monthly, origin-specific sampling weights. EPA did not have access to the underlying sampling weights and was unable to derive them from the count data provided.

Table A2 demonstrates a substantial discrepancy between the estimate of total people and the *actual* number of total people from the landings data between the two years in the comparison. The Provider 1 data exhibit a 900% decrease in people between 2019 and 2022, while the landings data show nearly a 50% increase in people. When EPA raised this discrepancy with Provider 1, they explained that their device sample underwent large changes in 2021 and 2022 due to disruptions in the location data market. Provider 1 proposed a number of meta-data corrections, such as making use of information about the total device activity seen across the vendor network. Although this was not officially communicated, EPA researchers inferred from the provider's comments that changes in the underlying apps providing location data (particularly Life360[v]) were likely a major factor in the discrepancy.

Ultimately, the EPA team abandoned the Provider 1 dataset for examining temporal trends. This demonstrates the importance of reference data, where a simple comparison with a reference dataset across time allowed EPA to identify an obvious flaw in the MD they obtained. The black-box nature of the provenance of the raw location data, coupled with the proprietary processing by vendors and fast-paced changes in the location data space places the responsibility on users to do basic validation on mobility datasets. This example illustrates that data vendors react to big movements in the underlying data supply and that reaction can negatively affect data quality.

## 3.2 California beaches - Data from Provider 2, Provider 4, and local agencies

In our second case study, we explore the suitability of MD for assessing recreation use impacts from an October 2, 2021 oil spill near Huntington Beach, CA. Around 25,000 gallons of oil were spilled from an offshore pipeline. A number of beach sites were fully or partially closed in the days following the spill. A key research question is how many beach trips were diverted from the affected coastline, both at closed and unclosed sites. And further, were there lingering impacts to visitation after closures were lifted due to oil and tarballs coming ashore for several additional weeks in Orange and San Diego Counties? In this section, we evaluate the suitability of MD from multiple vendors for this oil spill event-study style of analysis.

Before using MD to estimate the effects of the spill on recreation, we first consider the time series properties of the data, as was done in the Haines Alaska case study. We start with a within-dataset comparison of the Provider 2 MD over time, which was obtained from 2019 to 2022. Specifically, we evaluate how the Provider 2 counts correlate with explanatory variables like weather, day of the week, and holidays at two points in time - in 2019 and 2022. We would expect the relationships between the MD and exogenous weather variables to be relatively stable across time periods. Based on established patterns, we expect beach use to be higher on warmer days, weekends, and holidays, holding all else equal.

We use weather data for John Wayne International Airport (a representative location in Orange County) from NOAA's Climate Data Online portal[vi]. We use Provider 2 daily count data for four beach sites that were closed as a result of the October 2021 spill - Bolsa Chica State Beach, Huntington State Beach, Crystal Cove State Park, and Laguna Beach. We estimate a standard Poisson count model of the Provider 2 counts as a function of the maximum daily temperature, weekend/holiday, site dummies for each beach site, and a constant. The count models were estimated separately for 2019 and 2022 so the results could be compared across years.

Appendix Table A3 shows the results of these regressions. In 2019, the weather and weekend/holiday variables behave as expected, with use increasing on warmer days and on weekends/holidays. In 2022, the regression shows no significant relationship between temperature and use. This result is inconsistent with established understanding of recreational use patterns as one of the most important factors determining beach use is weather, and temperature specifically. This raises concerns about the stability of this dataset over time. This type of simple validation can help future researchers evaluate their mobility datasets.[vii]

Our next evaluation of MD for this context compares it to a high-quality reference dataset. As described in Section 2, we obtained visitation data collected by California State Parks. Reliable reference data of beach visitation based on paid vehicle receipts are available for Bolsa Chica State Beach and Huntington State Beach in Orange County.[viii] The vehicle counts at these sites are conducted with a robust and consistent methodology from staffed entrance kiosks. State Parks periodically uses site-specific studies to develop people-per-vehicle multipliers, which are constant over the time period of our analysis. Therefore, we simply use the paid vehicle counts in our analysis. In the appendix, Figures A3-A6 show that the AOIs used in MD queries were drawn to omit the parking areas, in addition to some ambiguous pieces of the general area, in all cases. For Bolsa Chica State Beach in particular, the vast majority of visitation is expected to be paid vehicle entrance, and walk-in visitation is unlikely to vary much or constitute a large share of the total. We present the daily Provider 2 MD in comparison to the daily paid vehicle counts at Bolsa Chica State Beach. The comparison at Huntington State Beach (not depicted) provides similar results.

<Figure 1 here>

Figure 1 shows this comparison for each day from 2019 through 2022, with separate panels for reference data, MD, and their ratio. We use 7-day moving averages of the counts to smooth out daily variability. In the first panel, the state park data demonstrate fairly stable seasonal patterns which hold across years. Visitation is highest in mid- to late-summer and is lowest in mid-winter. The MD show a broadly similar seasonal pattern in 2019 - 2021, but the pattern does not manifest in 2022.

We observe a spike in MD observations in the middle panel in April and May of 2022, also seen in the drop in the third panel from over 250 to below 10 in only a few days. After this spike, the following summer observations (e.g., July 2022) are significantly lower than the previous three years' summers, where visitation is typically highest. Further, July is the summer low-point in 2022, whereas it is the peak of summer observations in the other three years. This shows some clear ways in which the MD are diverging from the reliable reference data.

The third panel shows the calculated ratio of paid vehicles per unique MD count, which allows us to visualize how the relationship between the data sources evolves over time. It is possible that the relationship between the reference data and true visitation changes over time as well. Many of the same evaluative approaches we discuss in this paper also apply to reference data, though the conceptual model will generally be different. These observations support the following recommendations for initial evaluation steps for MD:

1) Identify patterns within years, compare patterns across years, and attempt to explain observed variation.
2) Compare any existing *a priori* expectations about true recreation patterns (e.g., known seasonal use patterns, construction closures, opening of new amenities) to MD and reference-data observations.

An example of step 1 above, applied to the reference data in Figure 1, shows consistent patterns across years. For the MD, we see significant divergence between the data in 2022 and the previous years. Recommended evaluation step 2 could make use of *a priori* expectations of typical seasonal patterns where the winter is the lowest use part of the year and summer is the busiest season. This is met by both datasets fairly well with the exception of 2022 in the MD again. In both datasets, we also expect to see the onset of the novel coronavirus (COVID) pandemic (March 2020) and the forementioned oil spill (October 2021). In the reference data, there is a data gap corresponding to the first few months of COVID, as expected given the policy at the time. In the MD, there appears to be a dip followed by an increase in summer visitation, which is also consistent with expectations about outdoor recreation in 2020. In both datasets, we see some basic visual evidence of a drop (relative to previous years) in visitation in early October 2021, corresponding to the oil spill.

In summary, we find both datasets in this evaluation to be potentially useful to answer questions about recreation, especially in the earlier years. The reference data appear to be consistent across years and conform to expectations (with respect to seasonality, COVID, and the oil spill). However, the divergence from both historical patterns and expectations remains a puzzle. These issues, combined with the April/May 2022 spike referenced above, cast doubt on the precision of MD for an event study of the 2021 oil spill, given its temporal proximity to the 2022 data anomaly. Identification of a spill effect requires consistency in the data before and after the spill.

Another perspective on MD performance in this context, shown in Appendix Figure A2, uses data from an additional vendor - Provider 4. We compare the mobility datasets from Providers 2 and 4 at Bolsa Chica State Beach for 2019 through 2024, though the available datasets only overlap in 2021 and 2022. For the period of overlap, the datasets from different vendors paint a different picture of use patterns, which raises questions about the reliability of both datasets. The analysis identified problems observed with using Provider 2's data as a proxy for visitation, and Figure A2 shows clear problems in the Provider 4 MD as well. Interestingly, the problems with Provider 4's data do not appear to correlate with the problems in Provider 2's data. Provider 4 data in 2021 and 2022 depict a lack of seasonality. Provider 4 in 2023 begins to show a seasonal pattern, but as we will show in Section 3.3, the Provider 4 data fail to track day-to-day MLB attendance data.

Our final evaluation of MD for this California analysis looks at the data around the time of the October 2021 spill for sites that were closed and known to have very little or zero use. These four sites are Bolsa Chica State Beach, Huntington State Beach, Crystal Cove State Park, and Laguna Beach. Closure information was compiled from publicly available news sources documenting the presence of closures. Non-recreators (e.g., cleanup crews or site management staff) may have represented some limited use of beaches during the closures, but this would pale in comparison to normal use levels. Figure 2 shows Provider 2 counts during the closures and surrounding weeks at these four sites. The dark gray shading represents a full closure (i.e., sand and water) and the light gray shading is a partial closure (i.e., water only).

<Figure 2 here>

The Provider 2 counts do not always demonstrate the expected pattern with respect to closures. The MD counts at Bolsa Chica decline during the closures, relative to pre-closure counts. Of the four sites, Bolsa Chica is generally the most isolated from external sources of MD (see Figure A3), and thus can be held to the highest standard with respect to the expected effect of a closure on visitation. The MD counts at Huntington State Beach and Crystal Cove may show reductions in use, but MD observations are still greater than zero. Weekday visitation in particular is in the same range as many non-closure period observations. Laguna is the site with the most notable divergence from expectations, with visitation patterns appearing to be unaffected by the closure. This could be explained by its proximity to a high-quality, non-beach park (which was not closed when the beach and water access closed) on the other side of the boardwalk (see Figure A6). The Provider 2 dataset does not apparently indicate a decline in visitation, yet these beaches were fenced off during the closures and not available for use.

The datasets with the most pronounced closure effects are generally the most isolated from irrelevant sources of MD. This highlights the importance of spatial error in MD analyses. The ability of MD to detect known, significant drops in visitation on a short time frame is a reasonable expectation that proves difficult to achieve. Duff et al. (2025) examine this objective in detail for a set of closed parks in Texas, finding significant variation in the strength of the closure effect across similar sites. In particular, the contribution of spatial error from surrounding roads, neighborhoods, and other sites of interest (such as the park next to Laguna Beach) can lead to significant measurement error. Our analysis of these four California parks gives further evidence of the importance of site characteristics in the context of MD analysis.

## 3.3 Evaluating mobility data using MLB stadium attendance data

In our final case study, we compare mobility data from multiple vendors to MLB stadium attendance data. MLB stadiums and attendance data offer several advantages for understanding the way that mobility data compare to actual attendance. These data provide numerous observations, as MLB teams host up to 81 regular season games per year in their home stadium, and teams report attendance for every game. MLB attendance data was unavailable for the 2020 season due to COVID. While stadiums are similar in size and shape, there are a few idiosyncratic differences that motivated our choice of a common AOI definition across stadiums. In the following analyses, we define the MD counts relative to a circle with a 125-meter radius surrounding second base (approximately the centroid of most stadiums; see appendix Section 8.4 and Figure A7 for discussion of this assumption). Using this setting as a backdrop for analysis, we demonstrate some notable patterns that warrant further investigation and comparison with other MD sources.

In nearly all MD research applications of which we are aware, the question of key importance is how many true visitors are to be inferred from a MD observation. We refer to this relationship as the “coverage rate,” the ratio between observed MD counts and true counts of visitation over any given time frame. The inverse of a coverage rate is referred to as the MD “expansion factor.” Although other functional relationships between MD and reference datasets have been found to exist in application (Merrill et al 2020, Tsai et al. 2023), we simplify the following discussion with an assumption of linearity.

<Figure 3 here>

Figure 3 shows a typical pattern in the calculation of MD expansion factors on a log scale at an example MLB stadium. At this particular stadium (American Family Field in Milwaukee, WI), expansion factors in 2019 ranged from around 25 to 100, with a similar range (aside from a few outlier days) in 2021. The MD analyzed from 2022 appears to perform so poorly as to be unusable. We observe a wide fluctuation in expansion factors and a gradual trend toward a very high average expansion factor.

Figure 4 suggests a simpler explanation for what went wrong. This figure shows MD counts for “game days” and counts on non-game days. Of particular interest is the mobility dataset’s ability to detect actual spikes in visitation on days when they really occur.

<Figure 4 here>

In 2019 and 2021, the MD counts on game days were systematically higher than counts on non-game days, suggesting that MD counts are capturing the effect of MLB attendance. There are occasionally non-game days with high attendance that may be attributable to other events such as concerts (for example, in early Oct 2021). Because we lack systematic data on these events, we assume that a few isolated occurrences are not evidence of a problem with MD.

However, 2022 demonstrates different patterns. Clustered groups of MLB games at a stadium (also known as homestands) are showing the correct number of visitation spikes in the Provider 2 data in 2022, but not on the dates that correspond to the actual attendance. Meanwhile, there are a corresponding number of game-day counts that are nearly zero. This is a misclassification pattern common across stadiums that suggests a back-end systematic problem with the timestamps of individual observations, rather than (for instance) the gradual accumulation of contribution from structural factors over time.

Beginning in July 2022, the MD fails to track any games at any of the stadiums, with a sharp decline to nearly zero for all observations. This example highlights the possibility that problems with mobility data could be caused by simple processing errors. The stadium data presented in this paper provide a setting that is well-targeted to identifying this particular problem: a location where there are events with visitation spikes that are orders of magnitude larger than the visitation on common non-event days (e.g., stadium staff). If this problem were to manifest in a setting with smoother daily visitation (e.g., public parks), it would be hard to tell if (for instance) a Monday's visitation were swapped with the same Thursday's visitation.

We can also analyze MLB stadium-years as separate datasets to explore spatial and temporal patterns in data quality. In our dataset of regular-season MLB games, a year maps cleanly to a season of professional baseball, generally April through October. A simple approach is to calculate the bivariate correlation between MD and attendance counts for all game days in a stadium-year. This correlation statistic relates to the stability of the coverage rate within a stadium-year. While MD with higher correlation to a reference dataset might not be holistically stronger than another, a strong correlation is at least an indication that changes in visitation are being reflected dynamically. As with all evaluation approaches in this section, the appropriateness of this comparison may depend on the specific application.

Appendix Table A4 shows the results of one such comparison across vendors and years. Each cell represents the mean stadium-level Pearson correlation coefficients between MD counts and MLB attendance counts for the 27 stadiums we observe[ix] in each corresponding year. The correlations were much stronger in 2019 and 2021 than in the following years, regardless of vendor. This evaluation approach allows the researcher to make rank-order comparisons across vendors and year. While the specific correlation-based metric described here should not be relied upon too heavily, the pattern it suggests is clear. The quality of MD deteriorated across multiple vendors after 2021.

<Figure 5 here>

Figure 5 shows Provider 3 and Provider 4 counts, respectively, for American Family Field in 2022 - 2024. These two datasets share a 2022 pattern that differs from that in Provider 2's dataset. For Providers 3 and 4, in the early months of 2022 before the MLB season began, there was a relatively low, stable level of MD counts on non-game days. After the 2022 season began, MD counts on non-game days increased considerably, typically hovering around the same level as on game days. After the season ended, MD counts fell back down to low levels, except for a few very high-use days, a pattern that is common across all of the stadiums (and thus unlikely to be explained by non-MLB events).

One potential explanation is temporal imprecision of recorded activities (intentional by the vendor or not), where the expected visitation spike on game days is diluted across other nearby days. Subsequent datasets that we evaluated with this same approach (Provider 3 in 2023 and 2024, Provider 3 in 2023) did not exhibit this specific pattern, but they did show a general lack of distinction between counts on game days and counts on non-game days.

The examples above demonstrate notable unexplained problems with mobility data. These findings should encourage other researchers to apply evaluation approaches like these to other vendors so that results can be compared. Given the usefulness of considering how these data are derived and the processes that govern the relationships between MD and true recreation, in Section 4, we propose a conceptual model of the mobility data generating process and discuss some likely reasons for the problems that we have demonstrated in Section 3.



# 4. Conceptual Model: Data Generating Process

Mobility datasets like the device location data used in this paper have a number of characteristics that are common across vendors and can present technical challenges for the researcher. *Ex ante*, it is difficult for a researcher to form realistic expectations about data quality based on the vendor's marketing claims. Vendors have the incentive to understate the importance of these destabilizing factors, while overstating their ability to mitigate them through proprietary techniques. Ultimately, the researcher must evaluate the finished product and decide whether it is of sufficient quality to support the desired analysis. In this section, we explain some of the ways in which a mobility dataset can be unstable across vendors, space, and especially time. Here, we focus on defining the issues and exploring how they relate to "known unknowns" in the data generating process.

One might assume that MD reflects a simple random process where each person in an AOI has a fixed probability of being sampled and therefore captured in a dataset. Yet, MD variability is not only a function of the number of people in the AOI, but also the variability in the process by which those people are identified in the MD metric. We describe a conceptual model of how several features of MD contribute to the relationship between true recreation and recreation as measured by MD. In this model, mobility datasets are decomposed into three categories: false negatives (which are not observed in the final product), true recreation, and false positives. The relative contribution of each piece in this process is of key importance to researchers - in particular how these contributions vary over time. Figure 7 is a stylized diagram illustrating the decomposition of recreators' recruitment into the MD sample.

<Figure 6 here>

This diagram is a decomposition of true recreation and a visual explanation of our conceptual model of the coverage rate. On the top left are true recreators, who may be excluded from the mobility data count if they do not satisfy each of the depicted conditions. We define "false positives" as observations that appear in the researcher's final mobility dataset, but are not actually indicative of a true recreation activity. We define "false negatives" as hypothetical observations of true recreation that do not appear in the final mobility dataset for one of several possible reasons. Each offramp box (e.g., "Device not with user") is a *mediator* which reduces the number of remaining recreators to be captured in the MD count. Each mediator is potentially influenced by a combination of exogenous (structural factors) and endogenous (decisions made by the device user) variables that influence the data generating process. Table 1 describes how each mediator is influenced by these variables, though this is not necessarily an exhaustive list.

<Table 1 here>

Even if a MD observation from a true recreator satisfies all of the mediator conditions for inclusion, data points can still be excluded during "black-box" processing steps undertaken by the MD vendor. For instance, vendors may decide to exclude an otherwise included data point if it is flagged as suspicious based on implausible location or timestamp patterns. Other data processing steps taken by vendors can lead to false inclusion. For example, many commercial data providers "salt" datasets sent to customers by adding a small amount of synthetic data for the purposes of anonymizing the data or identifying their product. All of the data processing done by vendors relies on different methods and assumptions, these methods differ by vendor and over time, and the methods are all proprietary and opaque to researchers who are buying the products (at least for every vendor that we are aware of including the ones that are used for this study). These details are generally articulated to the researcher in broad terms (e.g., "we use location data from a variety of apps in categories X, Y, and Z") and the full list is a trade secret.

In an attempt to account for the false negatives and false positives in Table 1, some vendors offer measures of aggregate visitation that extrapolate a total number of visitors from observed counts of mobile devices. For example, the Provider 1 data used in this study are accompanied by a proprietary “estimate” count, which extrapolates a total number of visitors from observed counts of mobile devices. This type of measure has the potential advantage of accounting for some of the variables mentioned above which can only be observed by the vendor, but it has the disadvantage of near-total opacity from the researcher’s perspective.

Because coverage rate is downstream of all of the factors described in this section, its stability is vulnerable. Some of these factors could be modeled using aggregate information to estimate the coverage rate (e.g., aggregate device ownership rates, timing of major OS changes, contribution of observable variables such as weather) but some potentially important factors are typically unobservable proprietary information (e.g., which apps contribute to a vendor’s supply). Sites of interest may have variable coverage rates because the mediators described above can vary according to visitor demographics as well. The temporal pattern of the contribution of these unobserved structural factors to the coverage rate is an open question. Some may have impacts that are smoothed over time (e.g., a contributing app slowly gains popularity), while others may have sharp impacts (e.g., an app suddenly discontinues a function that contributes location data). This dynamic is potentially important because it influences the comparability of MD estimates that are temporally proximate. Sharp changes in the coverage rate over time can invalidate comparisons even in the short run.

Although we do not have the metadata one might need to connect specific instances of MD instability to one or more root causes, the conceptual model provides a taxonomy of causes. Here, we apply this taxonomy to some of our datasets in this paper. Even without clear evidence of specific causal mechanisms, these potential causal categories can help provide context for the unusual or unexpected features of the data. In the California case study (Figure 1), we observe a sudden change in the relationship between MD and reference data in 2022. This change corresponded with a sudden supply shock from at least one important app supplying location data to multiple vendors (Life360, the same app that was likely affecting Provider 1’s data in the first case study). This may have caused a sharp drop in the number of true recreators who used a relevant app (Mediator 2) associated with the provider’s dataset. Meanwhile, Figure 2 and its discussion suggest a story about how MD visitation estimates at different beach sites are differentially impacted by spatial imprecision (Mediators 4 and 5), leading to noisy estimates during a time when true recreation was low. In the stadium data example, Figure 5 shows a breakdown in the MD counts' relationship with the game-day indicator variable. Although the timing of this problem is consistent with the same Mediator 2 explanation above, the pattern we observe with clustered game-day misclassifications implies an additional problem with the timestamps (Mediators 4 and 5). Other causes in this conceptual model might be demonstrated on other mobility datasets by analysis with respect to external data sources (e.g., industry data, timing of operating system changes, timing of privacy laws) or internal data sources (e.g., vendor-specific metadata, as described in the following section).

This conceptual model is our attempt to characterize some of the idiosyncratic complexities inherent to the mobility data generating process. The goal is to highlight the sources of uncertainty that upstream processing decisions impose on the researcher's understanding of their own analysis. In the discussion below, we suggest that mobility data vendors should provide researchers with supplementary datasets, metadata, and other insights they might have to enhance understanding of the mechanisms described in (or missing from) Table 1.



# 5. Discussion

This paper presents a number of empirical examples where MD fails to meet expected data quality standards. We illustrate examples where:

- Important patterns (e.g., seasonality) are missing, muted or degraded over time;
- Datasets from multiple vendors show incongruent results;
- MD fails to track with relevant conditions (e.g., closures and game days); and
- MD provides no apparently usable information at all.

The paper presents a conceptual model that describes structural features of the data generating process that may explain these issues. Researchers must carefully consider their assumptions about the coverage rate on a case-by-case basis. We recommend that researchers:

1) Examine new mobility datasets for evidence of instability and data supply problems. This paper shows several possible approaches for identifying problems if they exist. Demonstration of a positive outcome with respect to these approaches (e.g., 2019 and 2021 in Figure 4) can be presented as evidence of a strong mobility dataset.
2) Ask vendors to supply helpful metadata (see discussion below).
3) Choose analyses according to the strengths of the available dataset.
4) Report data weaknesses and caveats in publications.

A general diagnostic checklist for researchers with a new MD dataset is beyond the scope of this paper and current understanding, given the wide and multi-dimensional range of applications, MD products, vendors, and settings. Nevertheless, Section 3.2 provides a generic starting point for evaluation of a mobility dataset. Regardless of the endpoint, all outdoor recreation data queries can and should be examined for consistency of patterns across time (e.g., seasonality), and can be compared against externally-developed expectations (e.g., responsiveness to known disruptions).

In Section 3.3, we demonstrate a novel evaluation strategy applying MD to estimating MLB stadium attendance, which can be easily replicated with new datasets. Stadiums provide a rare set of nationwide reference data for MD validation, characterized by a sharply distinguished pattern of true visitation. Researchers should determine if their mobility dataset can precisely record peaks and troughs of visitation at the single-day level. Researchers should be creative in looking for additional datasets that can validate or ground-truth their MD dataset. Validation data is the key to expanding researcher's understanding of the MD, its weaknesses, and appropriate uses.

Mobility data may be appropriate for some recreation applications, but temporal instability will make certain approaches infeasible. Event studies aimed at detecting potential changes in recreation use are challenging to accomplish with currently available MD products. Temporal instability in MD, as this analysis demonstrates, can overwhelm the variation that a researcher is trying to detect. Calibrating the MD to reference data may not solve the problem if the underlying relationship is changing over a short time scale. To rigorously model the dynamics of the coverage rate, researchers will need more information about the data generating process than is currently possible. Simulating MD characteristic impacts on recreational demand models represents a productive direction for future research. Short of these more detailed approaches, estimates of changes in visitation will rely on strong assumptions about the stability of the MD coverage rate.

On the other hand, analysis aimed at understanding the relative visitation among sites might be a good application for MD. This approach may be largely shielded from the problem of intertemporal instability because the contribution of the exogenous variables to the coverage rate may simply be canceled out in the ratio. However, important considerations remain with this strategy. Researchers can define the spatial extent of their AOI in an unlimited number of ways, and the MD vendor's count estimate is directly influenced by an ultimately arbitrary boundary drawing. Matching the geographic extent between what the reference data are representing and what is sampled by the MD is an important task for making comparisons and MD-based visitation inferences. Challenges with defining boundaries exist more generally in recreation monitoring, but they are especially complex with MD.

Many applications for recreation using MD are less concerned with intertemporal comparisons and thus they are less concerned about temporal stability. For example, valuation models can exploit the home-origin information available from some MD providers. The patterns we describe in this paper might affect a valuation model to the extent that they skew the distribution of parameters of interest. We do not have strong expectations about how the pieces described in our conceptual model affect the demographic representation of the resulting mobility dataset. Future work on this topic should target the "secondary" variables included in some mobility data products, such as home origin locations and timestamps of observations.

One might hope that difference-in-difference approaches can be used to purge shared but unknown structural variability in MD. While difference-in-difference approaches might successfully account for the background trends in MD and observed covariates, we are typically interested in modeling or predicting the underlying quantities (e.g., true recreation) that MD is intended to measure, and not in the MD counts per se. This paper demonstrates that the relationship between observed MD and target recreators can change over time and space and be difficult to predict *a priori*. It would be optimistic to assume that an estimated difference in MD indicators at control and treatment sites that are each polluted with unknown variability would leave anything but a difference that is also polluted with unknown variability. This is yet another reason why untangling the structural variability in the measurement process from the measurement target is a critical next step in understanding how to best use MD data.

We hope that the ideas in this paper will motivate vendors to better communicate key background parameters with the research community. Examples include:

- Indices over time describing the data supply chain - number of contributing mobile apps per month, distribution of data volume contributed by individual apps (e.g., a Herfindahl–Hirschman type index), average number of apps by which a device is observed, distribution of volume by type of app, etc. This information could help researchers to diagnose problematic patterns (e.g., the 2022 divergence between data sources shown in Figure 1), identify eras of relative stability, understand changes in the data generating process, and optimize research designs accordingly.

- Indices over time describing the data sample – such as a constructed index of the number of unique devices observed within various levels of spatial aggregation[x]. This would be useful for helping to normalize the sample of visitors (using visitor origin data) at any given site and improve the researcher's ability to adjust for changes in representativeness (e.g., in Figures 4 and 5, changes in representativeness relative to other technical problems with timestamps). It would also be useful for identifying temporal shocks (common over space) in total data volume, which provides a useful control for contextualizing visitation changes in areas of interest.
- Documentation of the timeline of versioning of processing changes and samples, and how those changes affect data before and after the change. Rather than deliver a product with opaque ad-hoc adjustments, vendors should allow researchers access to the product upstream of these adjustments and provide the information required to assess them. This would allow researchers to better compare analyses and make a compelling case for robust results.

Additional studies are needed on the representativeness and temporal stability of MD datasets for popular applications, where strong assumptions are needed to create valid results given the problems documented here.



# 6. References

Csomós, György, Endre Márk Borza, and Jenő Zsolt Farkas. "Exploring park visitation trends during the Covid-19 pandemic in Hungary by using mobile device location data." Scientific Reports 13, no. 1 (2023): 11078.

Duff, Cameron, Christopher Giguere, and Jason Murray. "Remote sensing for quantifying recreation: An application to Texas public parks and NRDA" (*under review)* (2025)

Dutwin, David, and Trent D. Buskirk. "Apples to oranges or gala versus golden delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples." Public Opinion Quarterly 81, no. S1 (2017): 213-239.

Elliott, Graham, and Ulrich K. Müller. "Efficient tests for general persistent time variation in regression coefficients." The Review of Economic Studies 73, no. 4 (2006): 907-940.

Filazzola, Alessandro, Garland Xie, Kimberly Barrett, Andrea Dunn, Marc TJ Johnson, and James Scott MacIvor. "Using smartphone-GPS data to quantify human activity in green spaces." PLoS Computational Biology 18, no. 12 (2022): e1010725.

Goebel, Russell, Austin Schmaltz, Beth Ann Brackett, Spencer A. Wood, and Kimihiro Noguchi. "Modeling and forecasting percent changes in national park visitation using social media." Journal of Forecasting 42, no. 6 (2023): 1502-1518.

Gonyo, S.B. and Burkart, H. and Regan, S. "Leveraging big data for outdoor recreation management: A case study from the York river in Virginia" 354 (2024), https://doi.org/10.1016/j.jenvman.2024.120482

Hanson, Dieta, Emily J. Wilkins, Spencer A. Wood, Christian Crowley, Whitney Boone, and Rudy Schuster. 2025. Monitoring recreation on federally managed lands and waters—visitation estimation. Scientific Investigations Report 2025-5022.

Heckman, James J. "Sample selection bias as a specification error." Econometrica: Journal of the econometric society (1979): 153-161.

Lawson, Steve, Chris Monz, and Abbie Larkin. "Passive Mobile Data Analysis of Visitor Use in Parks and Protected Areas: Prospects and Challenges." Journal of Park & Recreation Administration 41, no. 4 (2023).

Leggett, C., Horsch, E., Smith, C., and Unsworth, R., 2017, "Estimating recreational visitation to federally-managed lands": Industrial Economics Inc., accessed July 15, 2024 at https://www.doi.gov/sites/doi.gov/files/uploads/final.task1_.report.2017.04.25.pdf

Liang, Yun, Junjun Yin, Bing Pan, Michael S. Lin, Lauren Miller, B. Derrick Taff, and Guangqing Chi. "Assessing the validity of mobile device data for estimating visitor demographics and visitation patterns in Yellowstone National Park." Journal of Environmental Management 317 (2022): 115410.

Merrill, Nathaniel H., Sarina F. Atkinson, Kate K. Mulvaney, Marisa J. Mazzotta, and Justin Bousquin. "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, no. 4 (2020): e0231863.

Merrill, Nathaniel H., Samantha G. Winder, Dieta R. Hanson, Spencer A. Wood, and Eric M. White. 2025. A national model for estimating United States public land visitation. Scientific Reports 15:42764.

Monz, Christopher, Milan Mitrovich, Ashley D'Antonio, and Abigail Sisneros-Kidd. "Using Mobile Device Data to Estimate Visitation in Parks and Protected Areas: An Example from the Nature Reserve of Orange County, California." Journal of Park & Recreation Administration 37, no. 4 (2019).

Parkinson, Colby, Bing Pan, Sophie A. Morris, William L. Rice, B. Derrick Taff, Guangqing Chi, and Peter Newman. "A Comparison of Tourists' Spatial–Temporal Behaviors Between Location-Based Service Data and Onsite GPS Tracks." Sustainability 17, no. 2 (2025): 391.

Song, Yang, Galen Newman, Xinke Huang, and Xinyue Ye. "Factors influencing long-term city park visitations for mid-sized US cities: A big data study using smartphone user mobility." Sustainable Cities and Society 80 (2022): 103815.

Struminskaya, Bella, Peter Lugtig, Vera Toepoel, Barry Schouten, Deirdre Giesen, and Ralph Dolmans. "Sharing data collected with smartphone sensors: willingness, participation, and nonparticipation bias." Public opinion quarterly 85, no. S1 (2021): 423-462.

Tsai, Wei-Lun, Nathaniel H. Merrill, Anne C. Neale, and Madeline Grupper. "Using cellular device location data to estimate visitation to public lands: Comparing device location data to US National Park Service's visitor use statistics." PloS one 18, no. 11 (2023): e0289922.

Winder, Samantha G., Spencer A. Wood, Matthew TJ Brownlee, and Emilia H. Lia. "Leveraging digital mobility data to estimate visitation in National Wildlife Refuges." Journal of Environmental Management 373 (2025): 123417.

Wood, Spencer A., Anne D. Guerry, Jessica M. Silver, and Martin Lacayo. 2013. Using social media to quantify nature-based tourism and recreation. Scientific Reports 3:2976.

Yabe, Takahiro, Massimiliano Luca, Kota Tsubouchi, Bruno Lepri, Marta C. Gonzalez, and Esteban Moro. "Enhancing human mobility research with open and standardized datasets." Nature Computational Science 4, no. 7 (2024): 469-472.

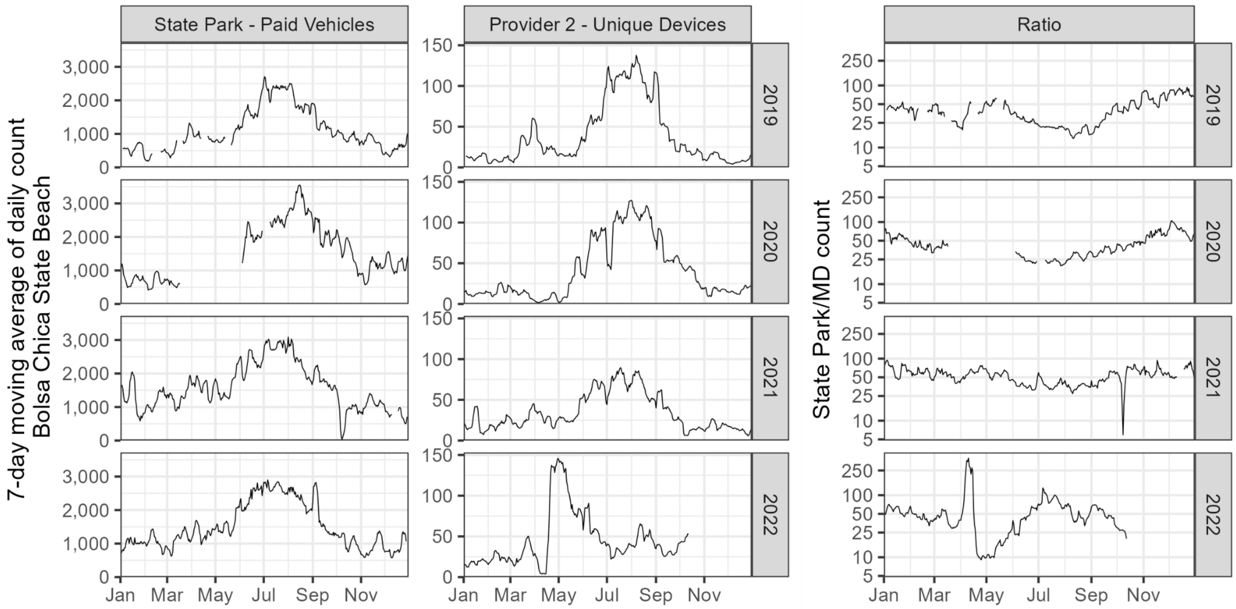
Zhang, Wendong, Xibo Wan, Wenran Fan, and Yongjie Ji. "Uncovering disparities in water-based outdoor recreation using cell phone mobility data." Environmental Research Letters 19, no. 11 (2024): 114057.

Table 1: Mediators and Variables - Conceptual Model

| Mediator | Exogenous variables | Endogenous variables |
|---|---|---|
| **False Negatives (Exclusions)** | | |
| 1: Device not with user | • Market for mobile devices | • Visitor demographics<br>• Site-specific factors<br>• Daily variables (e.g., weather) |
| 2: Relevant app not used | • Availability of apps in the market (e.g., Life360)<br>• Vendor contracts with apps | • Prevalence of relevant apps<br>• Intensity of use of relevant apps |
| 3: Location services not used | • OS changes (e.g., iOS 13)[xi]<br>• Changes in the way apps use location services | • Beliefs/preferences about location services |
| 4: Exclusion: imprecision | • OS changes<br>• Type of mobile coverage (e.g., Wi-Fi, GPS, 4g, 5g)<br>• Which app generates the location data<br>• Imprecision in digits of lat/lon, timestamps | |
| **False Positives (Inclusions)** | | |
| 5. Inclusion: imprecision | • OS changes<br>• Type of mobile coverage (e.g., Wi-Fi, GPS, 4g, 5g)<br>• Which app generates the location data<br>• Imprecision in digits of lat/lon, timestamps | |
| 6. Inclusions: other | • Fake data from apps<br>• “Salted” data points from providers | |

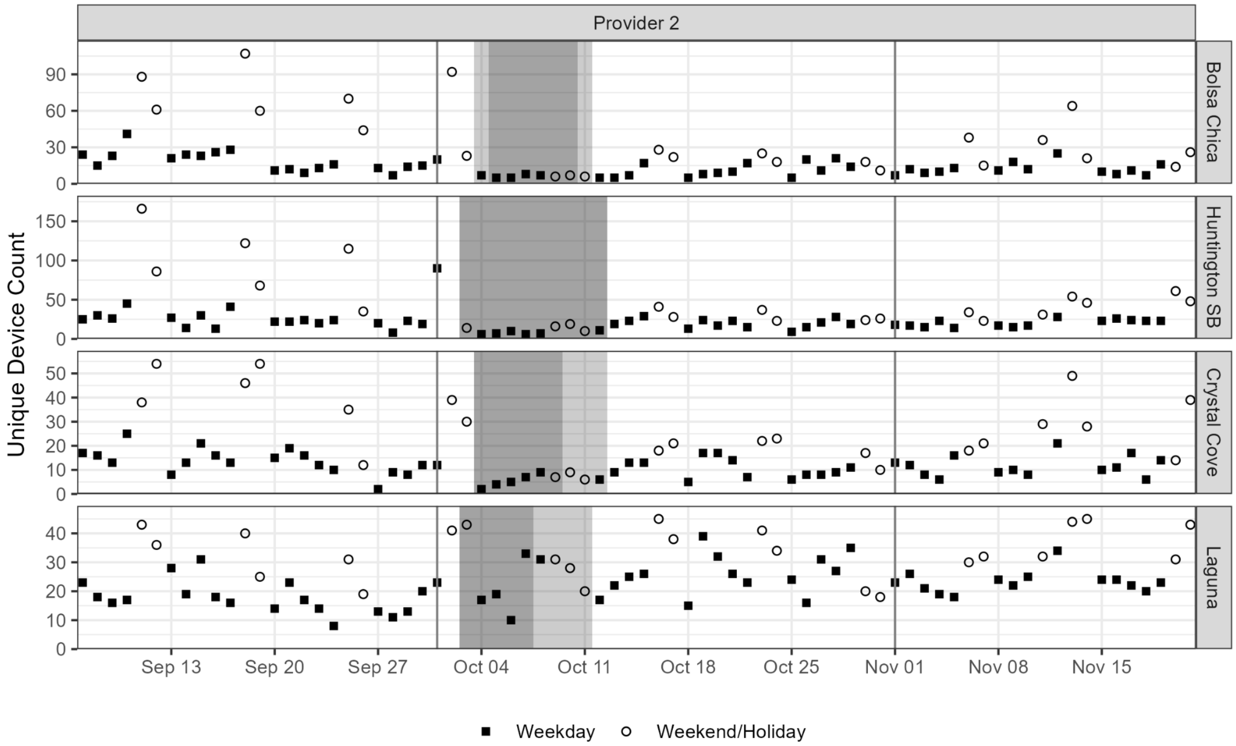
*Notes: This table connects directly to the diagram in Figure 6. Each bullet point is a potential explanation for why a “mediator” occurs. This list is non-exhaustive, and reflects our best understanding of the technology as it exists at the time of publication.*

Figure 1: Reference Data and Provider 2 - Counts and Ratio - Bolsa Chica State Beach



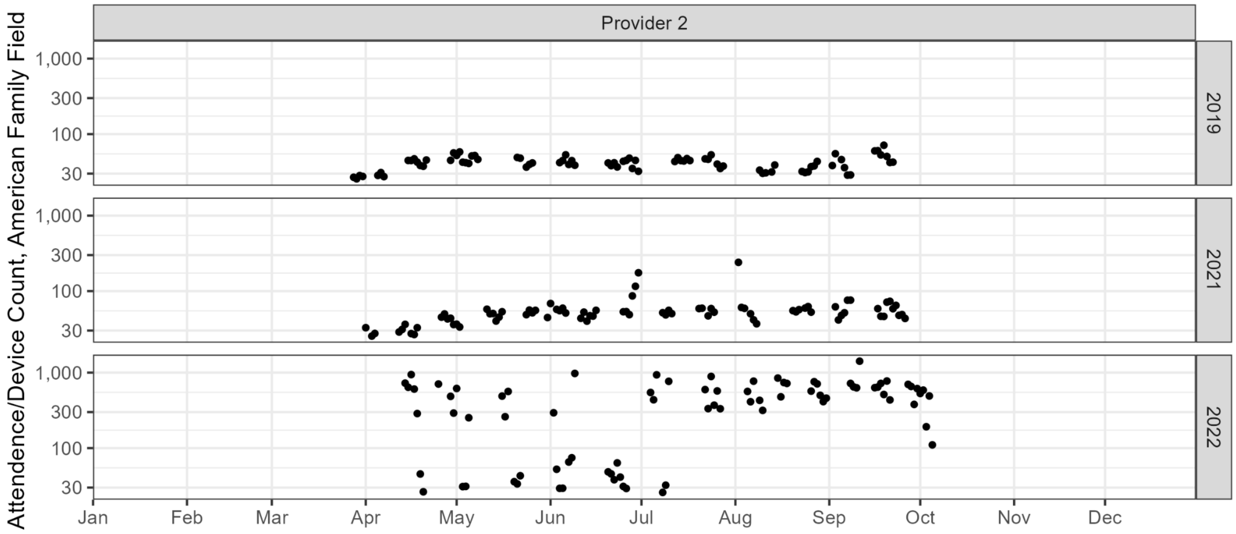

*Notes: We use paid vehicle entrants as a proxy measure for individual visits. The ratio in the third panel shows the number of paid vehicles (not unique visitors) per MD observation. The ratio is calculated relative to the 7-day moving average of both measures.*

Figure 2: Provider 2 - Counts During Closure Periods - Four CA Beach Sites



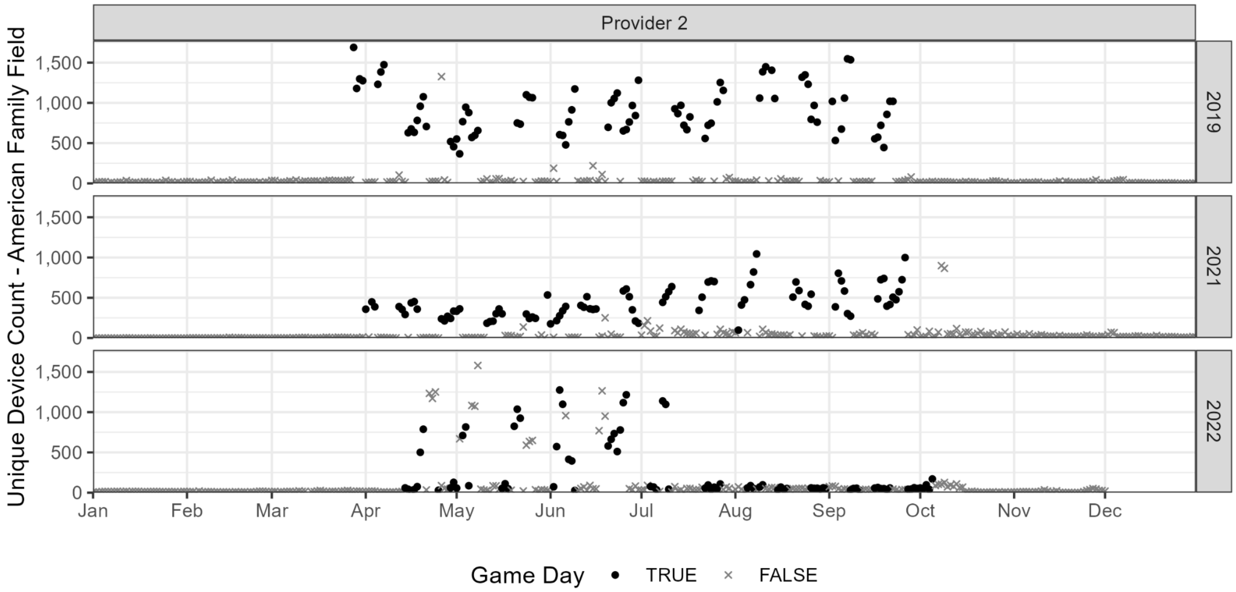

*Notes: See Appendix Figures A3-A6 for an image of the shapefiles used to generate these counts. The dark gray zone represents a full site closure, and light gray is a partial (e.g., water only) closure.*

Figure 3: Provider 2 and MLB Attendance - Daily Expansion Factors – American Family Field



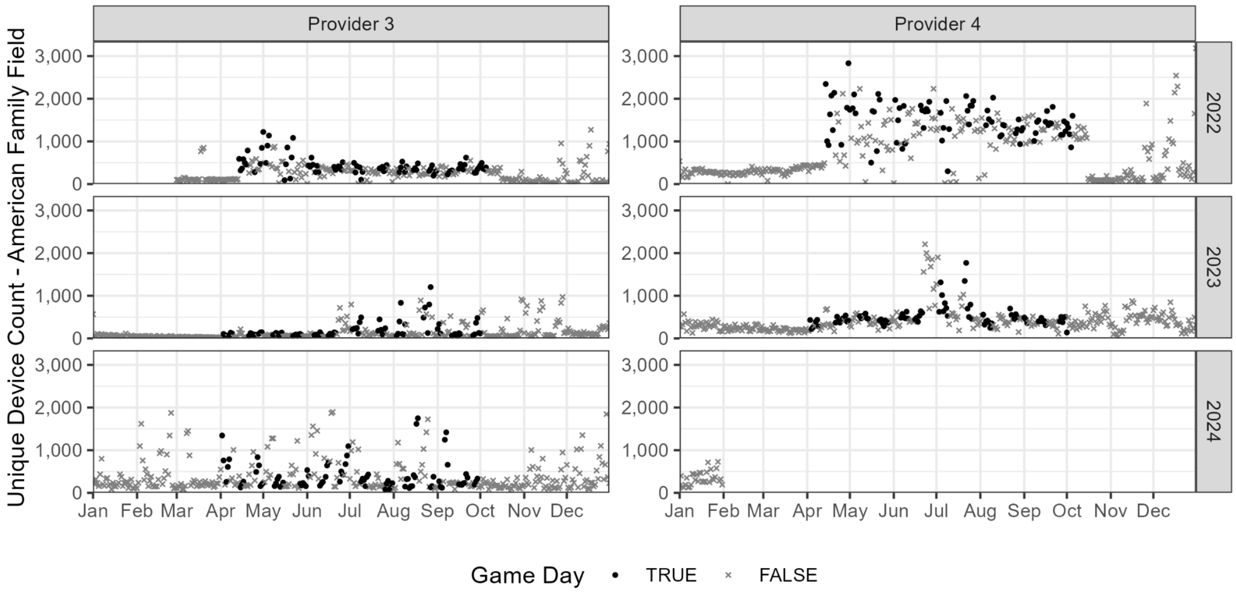

*Notes: This "expansion factor" measure is the inverse of the "coverage rate" measure. This is conceptually comparable to the third panel in Figure 1, but in terms of individual ticket instead of vehicles.*

Figure 4: Provider 2 - Counts on Game Days and Non-Game Days – American Family Field



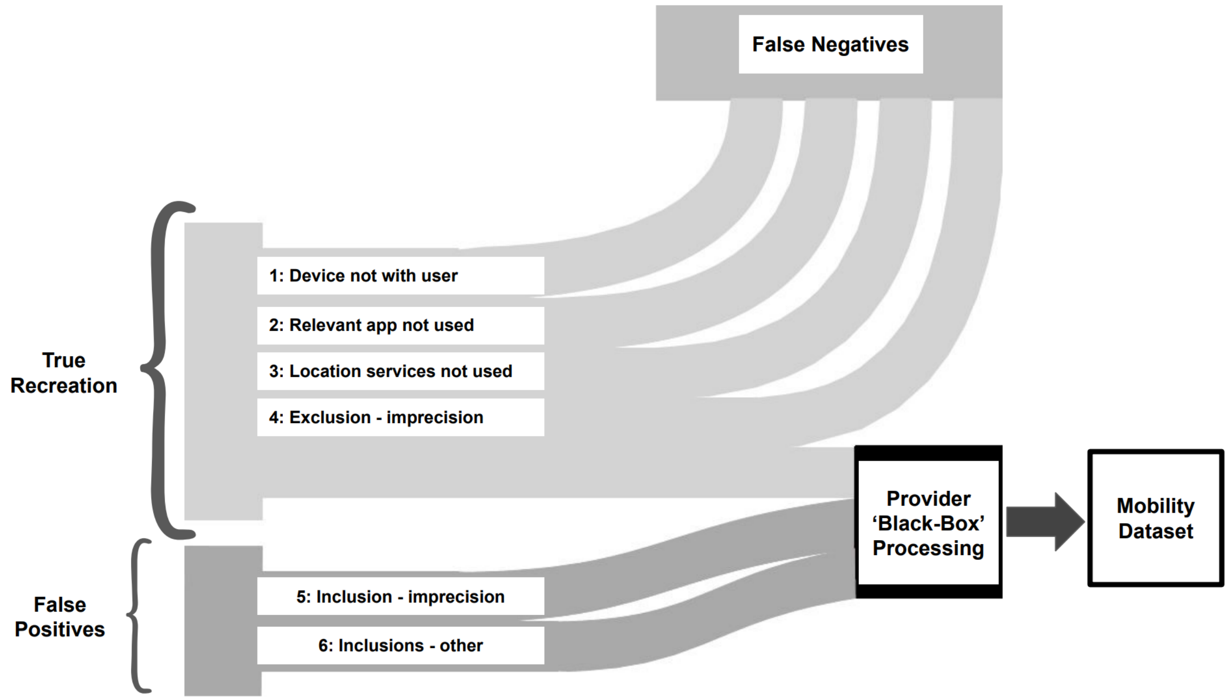

*Notes: This figure shows the raw MD counts associated with dates on which MLB games are scheduled. Instances of high counts on non-game days might be explained by non-MLB events using the stadium. Instances of low counts on game days (which become common in 2022) lack an obvious explanatory mechanism.*

Figure 5: Providers 3 and 4 – MD on Game Days and Non-Game Days – American Family Field




*Notes: This is a continuation of Figure 4 with two other vendors. For both datasets in 2022, systematically higher counts on non-game days coincide with the beginning of the season.*

Figure 6: Conceptual Model




*Notes: This chart shows a simplification of the mobility data generating process. True recreators either show up in the final dataset or are filtered out (false negatives) for a variety of reasons. Additional data (false positives) show up in the final dataset. The reasons shown in this diagram are likely non-exhaustive.*

---

[i] In this paper, we use the term mobility data to describe the class of data sources that are composed of observations of anonymized location-based information created by smartphones, watches, or other similar devices. Our analysis and discussion of these mobile device location data does not necessarily include other categories of data that are also described as mobility data: for example, georeferenced trip reviews and travel routes that are posted to online platforms. These data sources are comparable in some ways to the mobile device location data in this study, but there are conceptually different strengths and weaknesses to be considered.

[ii] We anonymize providers by using the "Provider *n*" moniker for each of the four mobility data sources used in the paper. Anonymization of the vendors was requested by multiple of the authors' employers.

[iii] The formats available reflect the timing of acquisition, generally within a month of the end of the corresponding time frame.

[iv] As discussed in Winder et al., (2025), datasets that were expected to be simply appended with new observations showed discrepancies in the overlap. This highlights the possibility that vendors may change their processing in a backward-looking way.

[v] Life360 was a major supplier of mobility data prior to January 2022. We cannot specifically connect it to the vendors we use for this paper, but examples like this are likely important for the stability of the data generating process: https://themarkup.org/privacy/2022/01/27/life360-says-it-will-stop-selling-precise-location-data

[vi] https://www.ncei.noaa.gov/cdo-web/

[vii] An extension would be to evaluate the robustness of patterns in different time periods at the same site. Mobility data offers the advantage of strong temporal coverage over a site of interest, and thus can often provide the number of observations required to estimate models with flexible interaction terms aimed at identifying more complicated patterns. For instance, if mobility data identifies that a site has higher use on alternating Wednesdays in the summer (e.g., a beach volleyball league), does this pattern hold up for other years when the same underlying mechanism should be driving an effect?

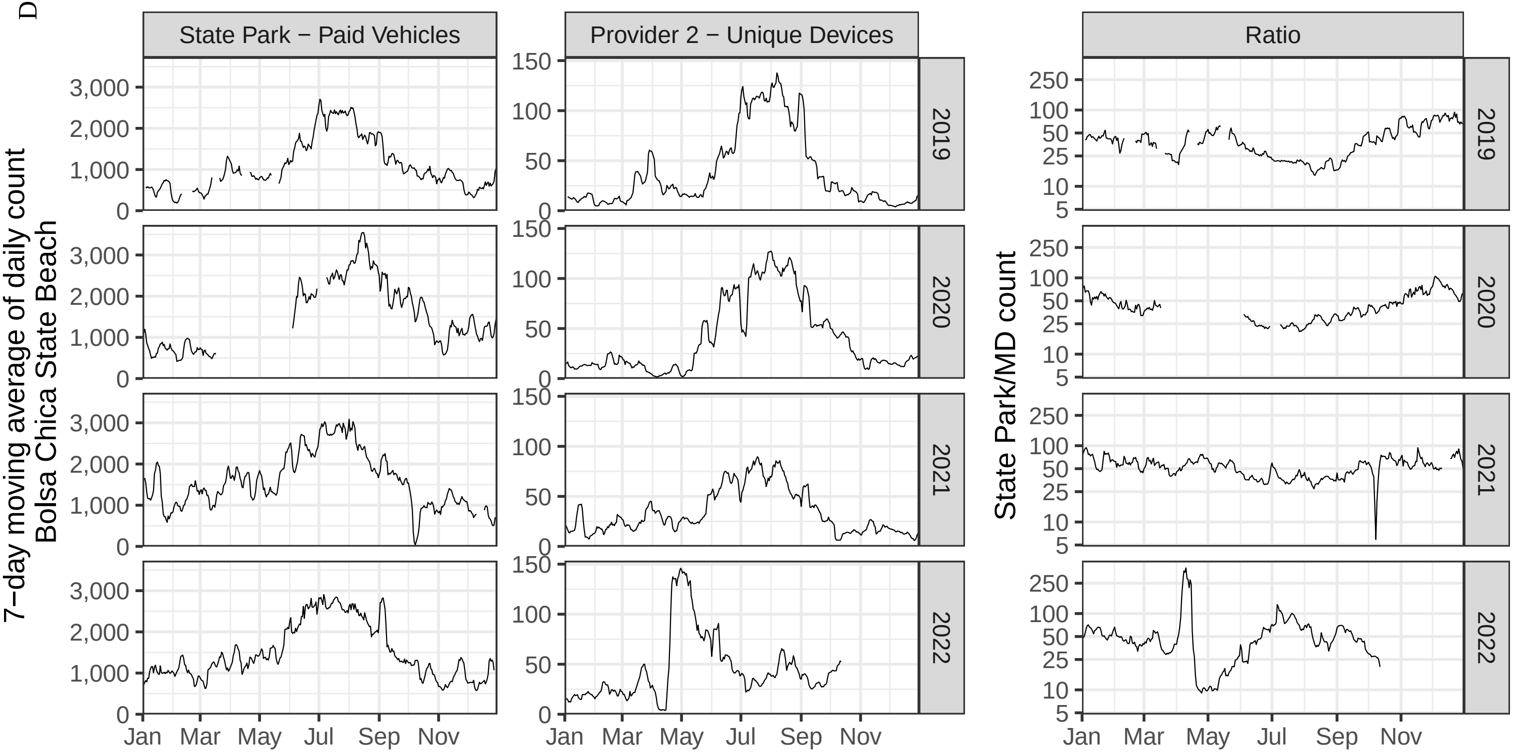
[viii] Data measures for walk-in use and free parking at these sites are considered by California State Parks to be less reliable, however, paid vehicle receipts represent the most significant share of use at these sites and are representative of the broader trends in visitation. While data for other state parks exist, there are reliability concerns due to low-quality counting methods, inconsistent data collection methods over time, a weak linkage to beach use (e.g., parking may support other non-beach uses), and other factors.

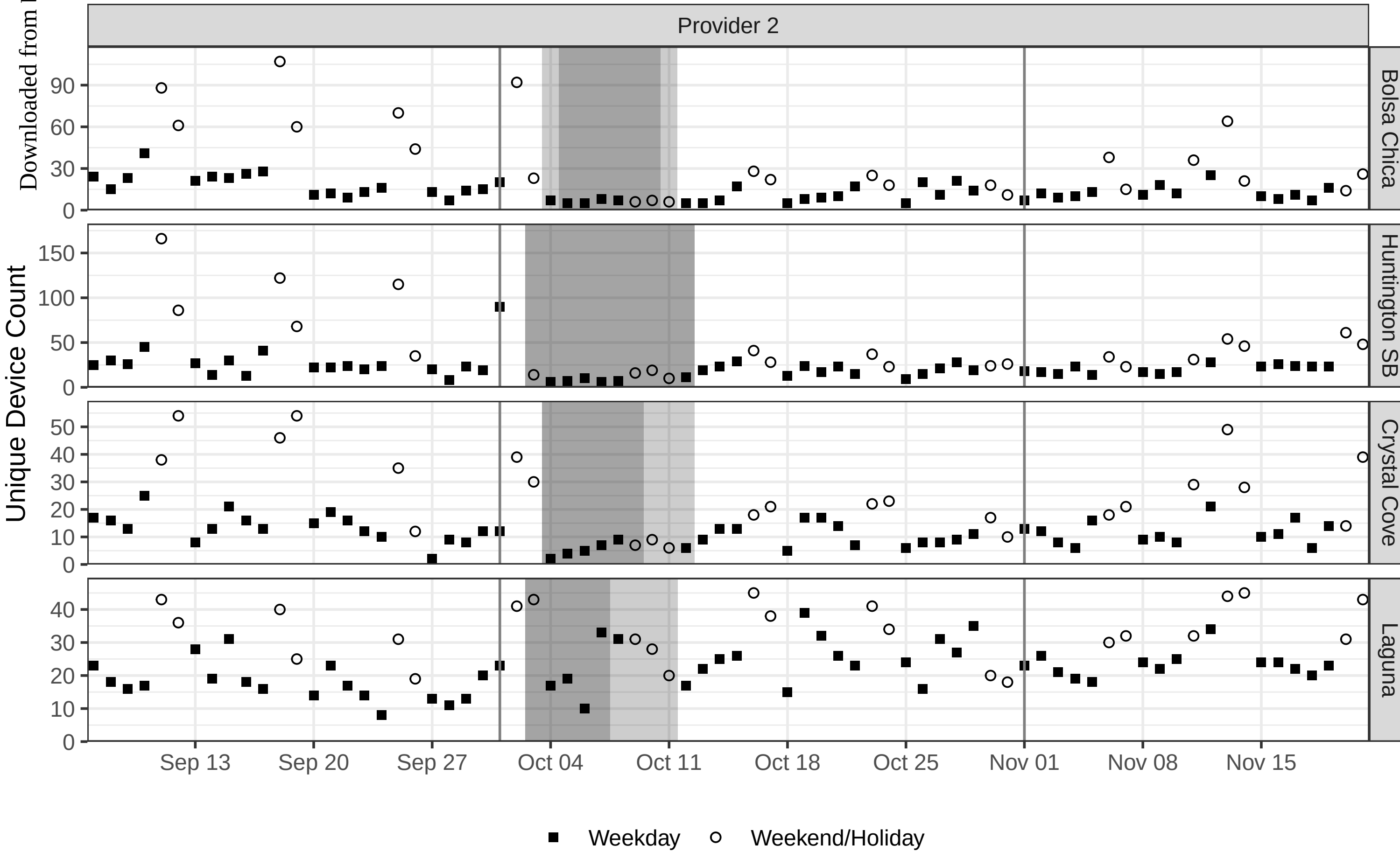
[ix] There are 30 MLB teams in total, but we do not analyze data from the stadiums in Toronto (mobility data vendors were US only), Arlington TX (team changed stadiums in 2020), or Detroit (site was too close to another stadium).

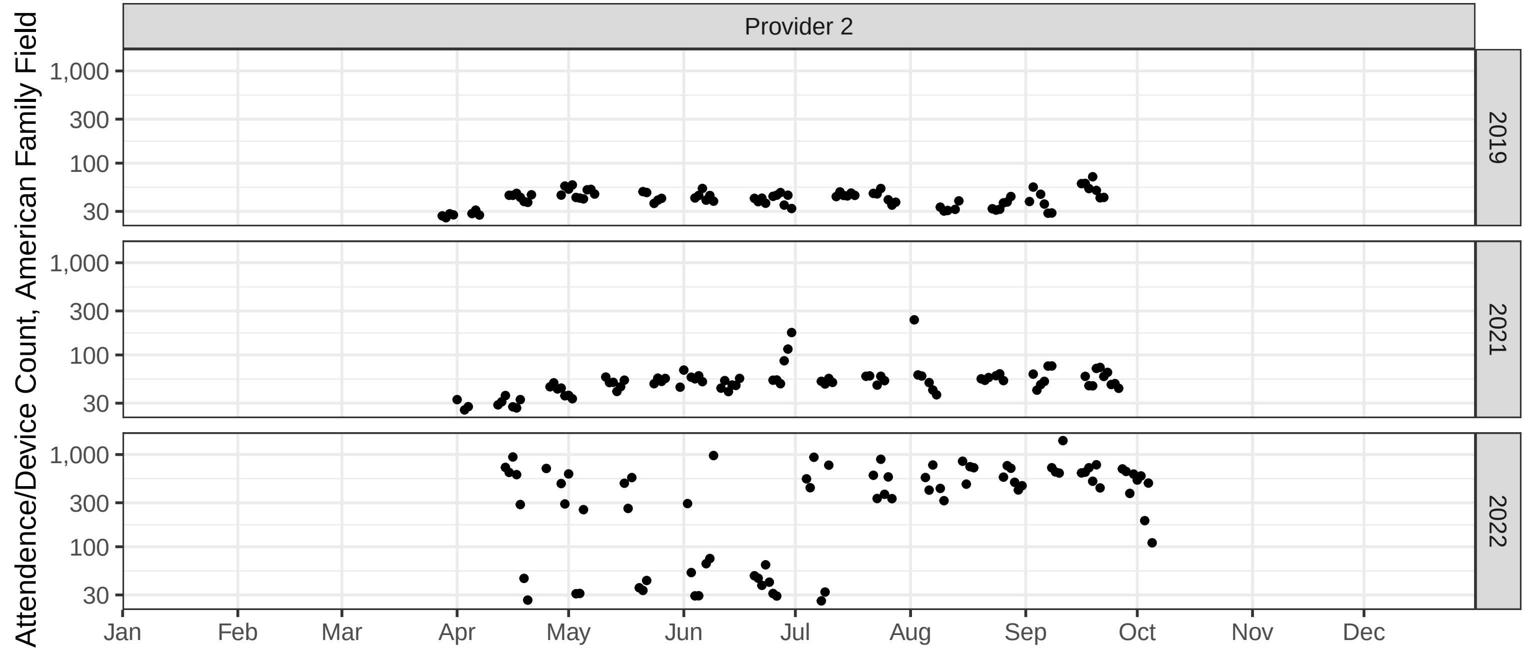
[x] We thank an anonymous referee for this suggestion.

[xi] iOS 13 made major changes in the way app users were prompted to allow use of location services, generally recording their location less often: https://bluedot.io/blog/location-prompt-ios-13-permissions/

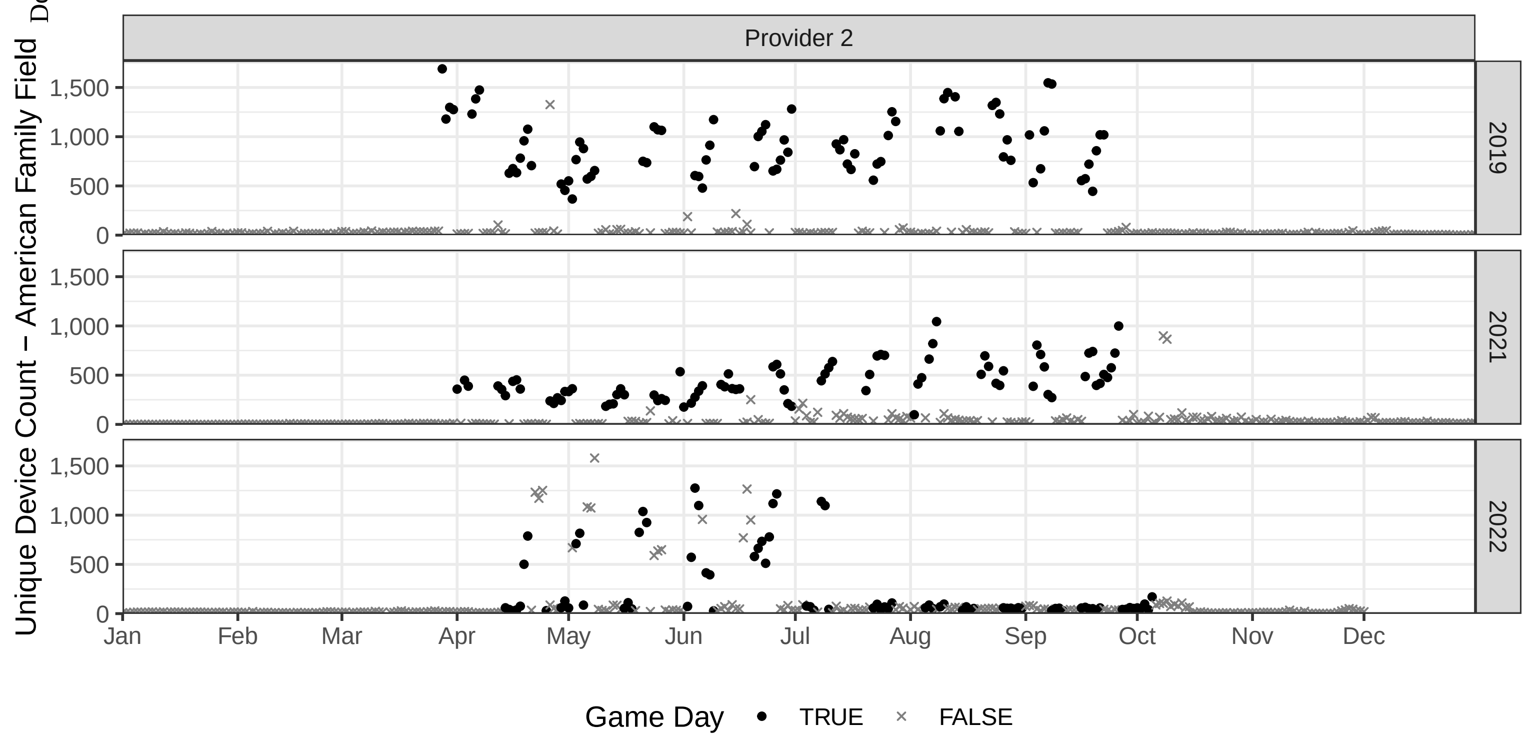
Do

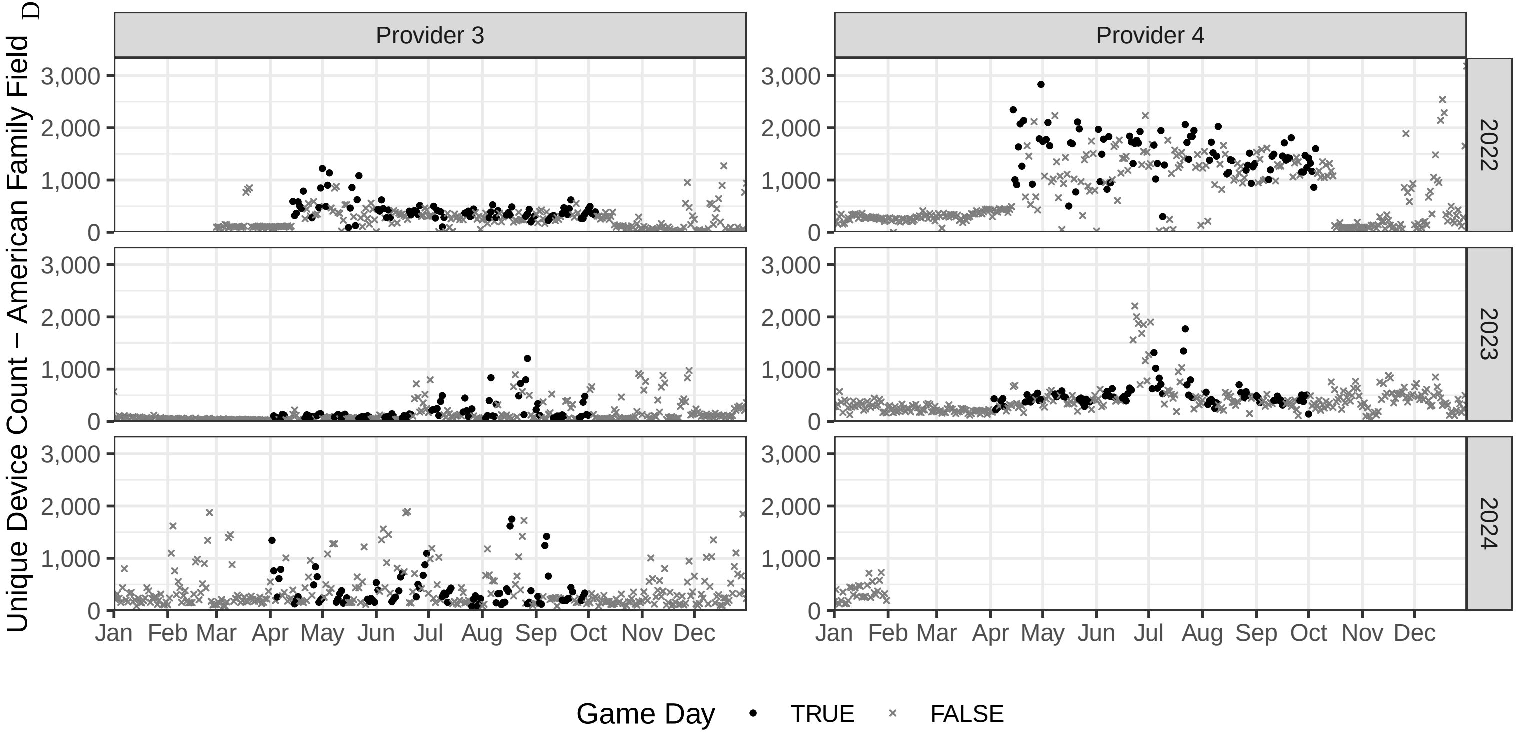

State Park − Paid Vehicles
Provider 2 − Unique Devices
Ratio
2019
2020
2021
2022
7−day moving average of daily count
Bolsa Chica State Beach
State Park/MD count
0
1,000
2,000
3,000
0
50
100
150
5
10
25
50
100
250
Jan
Mar
May
Jul
Sep
Nov


Provider 2
Bolsa Chica
Huntington SB
Crystal Cove
Laguna
Unique Device Count
0
30
60
90
0
50
100
150
0
10
20
30
40
50
0
10
20
30
40
Sep 13
Sep 20
Sep 27
Oct 04
Oct 11
Oct 18
Oct 25
Nov 01
Nov 08
Nov 15
Weekday
Weekend/Holiday

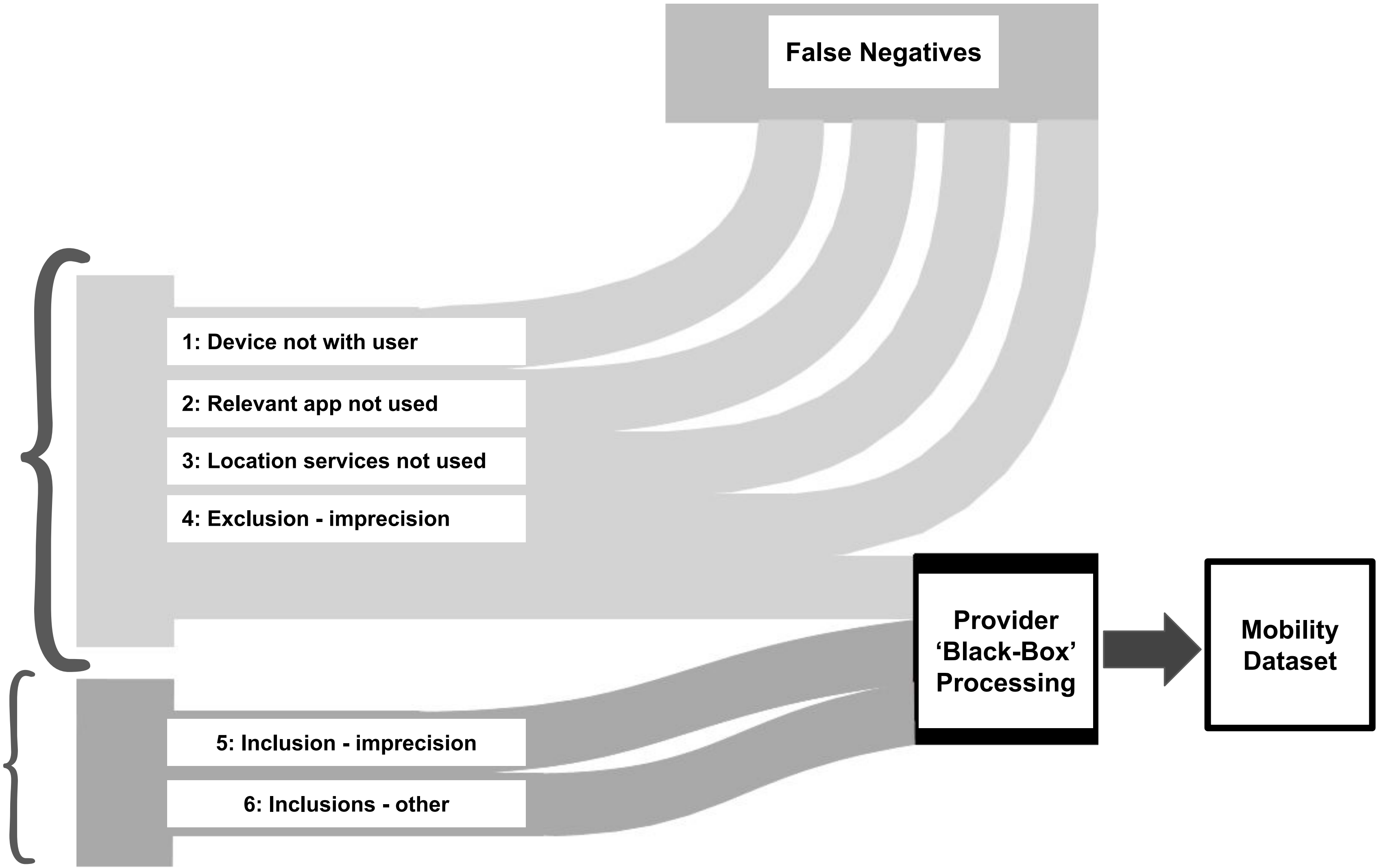


Provider 2
2019
2021
2022
1,000
300
100
30
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Attendence/Device Count, American Family Field


Provider 2
Unique Device Count – American Family Field
1,500
1,000
500
0
2019
2021
2022
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Game Day
TRUE
FALSE



Provider 3
Provider 4
2022
2023
2024
Unique Device Count – American Family Field
3,000
2,000
1,000
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Game Day
TRUE
FALSE


False Negatives
True
Recreation
1: Device not with user
2: Relevant app not used
3: Location services not used
4: Exclusion - imprecision
False
Positives
5: Inclusion - imprecision
6: Inclusions - other
Provider
'Black-Box'
Processing
Mobility
Dataset