Is a Photo Worth 1,000 Likes?
The Influence of Instagram at National Parks

Ashley C. Lowe Mackenzie*

Steven J. Dundas†

Abstract / Summary Paragraph

Recent increases in recreational visits, staffing shortages, and maintenance backlogs have left public land agencies in the United States (US) facing multiple stressors that affect their ability to manage land sustainably[1,2]. One supposition by traditional media outlets is that social media is the cause of recent increases in visits and resource degradation on public lands[3–7]. We test the validity of such claims and investigate how online engagement with content and user behavior on Instagram may connect to visitation trends at US National Parks. Using millions of georeferenced Instagram posts, we find a small, positive effect (5 to 7%) on visitation due to viral content and the presence of influencers at parks with moderate and high social media exposure. National parks with high exposure to Instagram see visitation increases from viral content regardless of who posts it, while parks in the moderate group see increases only from posts by influencers. We do not find evidence of a contemporaneous effect on visitation based on the volume of content posted to each park location. Our results suggest any causal connection between social media and visitation to public lands is likely through a behavioral channel, such as herd behavior or the bandwagon effect.

---

* Corresponding Author. Assistant Professor. Department of Natural Resources and Environmental Management University of Hawai'i at Mānoa, Honolulu, HI 96822, USA. Email: alowemac@hawaii.edu

† Associate Professor. Department of Applied Economics and the Coastal Oregon Marine Experiment Station Oregon State University, Corvallis, OR 97331, USA. Email: steven.dundas@oregonstate.edu

MAIN

Online activity often has offline consequences. Recent studies show that the dissemination and use of information from social media platforms has potentially significant implications for a wide range of societal outcomes, including democratic election[8–10], mental health[11–13], cognitive comprehension[14–16] and public health[17–19]. An additional focal point where claims of impacts from increased social media exposure proliferate is how people interact with public goods, especially parks and wilderness areas. Public lands are estimated to receive between 5 to 18 billion visits globally per year, with over 40% of that visitation occurring in North America alone[20]. Managing agencies of the nearly 640 million acres of public land in the United States (U.S.) have multiple objectives, including balancing the conservation of species and ecosystems with recreational use. It is this balance that is currently being tested by large increases in visitation and subsequent resource degradation. Recent research in the U.S. suggests the demand at certain parks may be influenced by online activity but the mechanisms by which social media impacts park visitation have yet to be clearly identified[21,22].

U.S. National Parks (USNPs) are often referred to as the "crown jewels" of the nation's public lands, requiring congressional approval for designation and represent areas that hold nationally significant natural, cultural and recreational resources. Trends in recreational visits to USNPs have experienced multiple periods of growth over the last century (Fig. 1). Steady increases in visitation following the end of World War II[23,24] were attributed to population growth and rising levels of income, along with improvements in travel and accessibility (e.g., development of the interstate highway network)[23–25]. In the early 2000s visitation trends were stagnant, leading researchers to examine the relationship to events like the 9/11 terrorist attacks or macroeconomic trends[26,27]. This period of stagnating interest was short-lived and USNPs began reporting record-breaking annual visitation beginning in 2013. Increased visitors adds stress to often under-resourced staff and aging infrastructure, while also contributing to adverse effects to fragile ecosystems[1,3]. To highlight the magnitude of the problem, consider that the deferred maintenance and repair backlog for the National Park Service (NPS) was $23.3 billion (nearly 60% of the total for all U.S. federal land management agencies) as of July 2024[2].

A common culprit described in traditional media outlets and land management networks for this recent rise in visitors and resource degradation in USNPs is the social media platform Instagram[4–7]. Instagram began as a photo-sharing application and has shouldered a significant amount of blame for the growing crowds for multiple reasons, including the timing of the app's release coinciding with the shift in visitation trends, the ability of users to georeferenced content posted withn the app, the way content is delivered to its 1 billion active monthly users and accounts by NPS staff reporting swarms of visitors observed taking self-portraits, or selfies, at particular scenic locations[6,7,21]. Perhaps more importantly, Instagram provides significant financial incentives for influencers, or popular app users, to generate content in places with unique and scenic landscapes to develop and monetize a brand associated with nature and travel. Investigation of the influence of Instagram on visitation to USNPs and the potential mechanisms driving in-app engagement may reveal important insights to improving sustainable management of these iconic places.

Data from social media platforms have been shown to be a relatively reliable proxy to estimate the number of visitors to public lands[28–32], suggesting avenues to reduce costs of data collection and extend visitor use estimates to wilderness areas where prior data collection was limited[28,30,33–36]. A recent study suggests that USNP locations which created online accounts on sites such Twitter and Instagram that receive high exposure to users have seen increases in visitation[22]. Linking elements of user behavior on social media platforms is noted as a significant gap in our understanding of the problem,[37] although recent regional evidence suggests certain locations are more susceptible to social media influence through high online engagement (e.g., large amounts of *likes*) and not the general act of sharing of photos of the location online.[21]

Considering how the potential influence of Instagram is driven by activity from the user base within the app (Fig. 2), we posit two hypotheses about potential links between visitation increases and Instagram. Higher visitation could be driven by 1) a reduction in search and information costs[38] provided by Instagram in helping people find recreational opportunities or 2) social media could be incentivizing herd behavior, or a bandwagon effect, whereas an individual's demand for a commodity is increased due to the fact that others are also consuming the same good[39]. In this paper, we test these two hypotheses empirically. The reduced search cost hypothesis is investigated using user data on the volume of Instagram activity georeferenced to a USNP and the bandwagon effect hypothesis is examined by identifying content with high user engagement and quantifying frequent user behavior indicative of monetized influencer accounts.

Here we examined over 8.3 million public posts to Instagram from 2010 to 2019 that are georeferenced to USNP locations with a varied set of analytical tools, including textual and sentiment analysis, machine learning, and econometric visitation models (Methods). We define the scope of our analysis to the 40 USNPs located in the conterminous U.S. that are accessible by vehicle for the average visitor and those that existed as a National Park prior to launch of Instagram in 2010 (Extended Data Fig.1). In these analyses, we leverage the timing of Instagram's rise to over a billion monthly active users and a social media exposure metric similar in spirit to Wichman (2024) but using different drivers from the app's user base (engagement, presence of influencers) to construct the index. We then use park-specific, contemporaneous upload activity and user engagement with content, and the information provided within this content posted to Instagram to operationalize each method. The size and scope of the available social media data and the empirical models applied lead to a set of results that suggest factors that drive online engagement, and how content with varied levels of engagement and different types of users of Instagram may have influenced visitor trends to these well-known locations.

We first estimate a count data regression model using variables generated from text data in captions of each post and posting frequency from each user to determine what types of information and in-app actions are associated with increased engagement with Instagram content. The inclusion of factors like points of interest and both positive and negative sentiment marginally increase online engagement, with the primary driver being user posting intensity. In other words, users who post content about national parks to their followers frequently (i.e., likely monetized influencer accounts) tend to see significantly larger engagement with their posts. An additional approach using random forest analysis confirms that user posting intensity is the primary factor driving engagement with USNP content on Instagram. In terms of visitation, our regression model span both pre- and post-Instagram time periods from 1993 to 2019. Looking at the broad impact of Instagram across parks based on social media exposure suggests no general increase in visitation associated with parks with high exposure, counter to results from Wichman (2024) using a similar instrumental variable approach to address endogeneity between visitation and social media exposure. Our preferred visitation model uses new social media exposure groupings and insights from textual, sentiment, and machine learning analyses on engagement and user characteristics. These results show that the volume of content posted is not associated with increased visitation. This suggests that we may reject our first hypothesis that social media reduces search and information costs and therefore may increase visitation. For the second hypothesis, we estimate that USNP locations that had "viral moments," where large numbers of users interacted with georeferenced content, saw small increases (~5%) attributable to the cumulative impact of all current and past park-specific viral content. We also show visitation to parks with the highest social media exposure were likely impacted by viral content posted by any user type while a second group of parks with moderate exposure experienced increased visits attributable to posts from high frequency users (influencers). These results provide important suggestive evidence on mechanisms for the link between social media and USNP visitation that is important when designing future policies for both social media use and public land management.

Influencers Drive Online Engagement on Instagram

Engagement online with Instagram has potential to influence offline behavior of users (Fig. 2). To understand what factors drive online engagement with posts georeferenced to USNP sites, we start with textual analysis[40,41]. We processed all text within the content of each of the 8.3 million geotagged Instagram posts in our dataset. For each USNP, we built park-specific dictionaries to represent the information in the text captions of all content. The dictionaries included identifying the presence of general location information (e.g., the state the park is located or the park name), points-of-interest listed by USNPs managers as important locations within the park and the activities listed as advertised *Things-To-Do* on each individual park's website[42] (Extended Data Fig. 2). We considered the extent of information provided by controlling for both the number of words within the caption and the number of hashtags (i.e., text preceded by the # symbol used to categorize content on social media). We also conducted a sentiment analysis to control for positive, neutral or negative language expressed in each post (Methods). The subjectivity of the each caption was analyized to capture whether the content contained more opinions, emotions, and personal interpretations or more objective and factual information.

We identify 2.8 million unique users posting content geotagged to USNP units and classify these users by summing the number of times an individual uploads content to any National Park geotag over our sample frame. We group users of the app by quantile, defining those posting content in the 90% percentile as “avid users”. This user type is of interest to our analysis for a few reasons. The ability to for content to go viral has led to monetization potential for Instagram account holders and those that are successful in this space are known as influencers. We do not directly observe if a specific user could be defined as an influencer or if they have a monetized account, but our classification of avid users is likely to contain many influencers who have large audiences on Instagram and attempt to sell lifestyle products associated with outdoor recreation and USNPs.

We used the above information in a negative binomial regression model to determine how these elements affect the count of likes each Instagram post is likely to receive from the general user base (Methods). We found user posting intensity had the largest effect on the amount of engagement recieved from online users (Fig 3.a). These avid users or influencers are expected to have an engagement rate 4.5 times greater than all other users. This result provides supportive evidence that the mechanism driving engagement is potentially more of a behavioral channel (e.g., bandwagon effect) than an information effect. That said, informational content also contributes to an increase in engagement as the presence of park-specific points-of-interest increased engagement by 1.1 times more than content from users not discussing specific locational information. The results also show that the number of hashtags in a post had a nonlinear effect, decreasing engagement at an increasing rate as the number of hashtags increased, while the number of words used within the captions had a positive effect. The number of likes increased if the caption expressed a positive or negative sentiment compared to captions which remained neutral (Extended Data Table 1). Lastly, we show that opinion-based information also had a positive effect on engagement relative to factual information.

To validate these findings, we use a random forest model to estimate the relative importance of each covariate in our model (Fig 3.b). Results from feature importance scoring found that user posting intensity is the most significant feature in understanding the level of engagement content recieves along with a strong postive relationship highlighed in path dependency plots (Extended Data Fig. 3).

No General Effect of Instagram on National Park Visitation

Our interest turns to visitation and the extent viral content and avid users may influence visitation to USNPs. We first show how results from prior work on social media's potential impacts on park visitation[22] compare with our finding that there is no broad statistically significant effect on visitation from Instagram generally (Fig 3.c). We find that the statistical significance of the large visitation effects found in Wichman (2024) is sensitive to how standard errors are clustered. The two-way cluster approach that includes clustering by year-month and by park-month implies that each park-month combination is an independent cluster, which obscures the natural structure of the data where correlations are likely higher within parks across time and could introduce excessive overlap and redundancy since the month component is represented in both levels[43]. We find that the statistically significance of this prior result disappears with a simpler and arguably more defensible approach to clustering the standard errors based on recent findings in this literature (Extended Data Table 2)[43–46]. Furthermore, our visitation model includes additional relevant controls omitted in this prior work, including population, demographic factors (esp. population over 65+, Fig. 1), trends in gasoline prices and fees. These additional variables are common in visitation models (Fig. 3.d) and their inclusion does impact the overall magnitude of the estimated relationship of interest and suggests a null result on the broad impacts of Instagram (Fig. 3.c, green estimates)[21,25–27,47–50]. Results suggest other factors affecting visitation include weather at each park, state population growth, the unemployment level and the changes in the population over 65 years of age[3,25].

Despite not finding a broad effect of Instagram on USNP visitation, the richness of our data allows us to explore our two hypotheses directly. To evaluate this in our context and create data-driven social media exposure groups among parks, we plot each USNP unit based on the number of likes and the sum of avid users in the 90th percentile (influencers) observed at each park. Using a density-based clustering method (DBSCAN), Yosemite and Zion NPs are identified as outliers. The remaining clusters are denser and thus we use a k-median clustering technique that is sensitive to outliers for the remainder of the analysis. The output from these processes suggests three clear levels of social media exposure: High (2 parks: Yosemite, Zion), Moderate (13 parks) and Low (25 parks) (Fig. 4). To show visitation trends among parks in our grouping strategy, we plot the average annual visitation for each park group across four decades (Fig. 5). The plot shows that parks in both the high and moderate groups saw an increase in average visitation shortly after Instagram began to gain potential influence in 2012[21], while visitation at the parks in the low exposure group appears relatively constant across time and the rise of Instagram. This suggests further exploration into what role, if any, Instagram may play in this increase in the moderate and high exposure groups.

Viral Content and Presence of Influencers Drives Small Visitation Increases

To investigate the information hypothesis, we specify a visitation model with park-specific data that captured monthly counts of georeferenced photo-user-days (PUDs)[28] representing the volume of daily user activity associated with each park per month on the app. We found the number of PUDs does not affect overall visitation at parks or within any exposure group, with the exception being a marginally significant positive effect in the moderate exposure parks (Fig. 6.a).

How Instagram content is shared to its users (Fig. 2) are important aspects of the functionality of the app and likely impacts how in-app behavior could translate to offline visitation changes, allowing a test of our second hypothesis (bandwagon effect). Social media sites have given users the ability to post content that can go viral, or spread quickly and widely online to large amounts of other app users. The process of going viral increases visibility for an individual's account and content. Social media sites such as YouTube, TikTok, X (Twitter) and Instagram have social networking paths that allow for popular content to expand its reach to larger audiences. Generally, the number of likes and comments an Instagram post received can be used to quantify the viral nature of content. Such content may affect visitation to USNPs differently than a contemporaneous measure of the count of georeferenced PUDs. Viral content online tends to be sticky. In other words, Instagram pushes content that receives high initial engagement to its users based on their social network, behavior online and location, increasing chances of a post reaching viral status and remaining visible to app users for longer periods of time.

We then include cumulative counts of content with high engagement as the aggregate of both past and current viral content and separate this content between influencers and normal users of the app. For each exposure group, we use posts in the 90th and 95th percentiles each year by park in terms of engagement (number of likes) and the sum of avid users posting at each USNP

(Extended Data Table 3). We found evidence that the cumulative viral effect from engaging content increased visitation at the two parks in the high exposure group and 13 parks in the moderate exposure group by approximately 5% per month (Fig. 6.b). Separating the effect from influencers, or avid users, and estimating the cumulative impact of their content leads to a slightly higher effect on visitation, up to 7% per month for high and moderate exposure parks. These results suggest that avid users had a larger impact than engaging content. As anticipated, neither the presence of influencers nor engaging content had any significant effect on visitation at the 25 parks in our low exposure group.

Discussion

Online activity has offline impacts[8–11,18,19,21,22] and visitation to USNPs is no exception. New information from social media has the potential to make the great outdoors less exclusive and increase the diversity of recreators[51] yet the conversation has often focused largely on its negative role in driving overcrowding and resource degradation. Our analyses that combine multiple modeling frameworks with data from millions of social media posts suggest that it is not the mere existence of Instagram that may drive increases in visitation but specific content and user behavior may be attributable to small increases in visits. At the park level, the reason for the increase attributable to Instagram may vary depending on the level of engagement of the user base with content georeferenced to that park and the presence of influencers.

Importantly, our analysis shows that the introduction of Instagram or the proliferation of geotagged USNP content alone (i.e., reduction in search costs hypothesis) does not explain the overall rise in visitation observed. We found that demographic trends also played a large role in the recent increases in visitation. For example, the national population over the age of 65 increased rapidly over this same time period (i.e., baby boomer generation reaching retirement age) and model results suggest a positive relationship with USNP visitation trends (Figs. 1, 3.d). This older demographic is more likely to be retired with increased leisure time to make a trip to a National Park. Other factors not directly addressed in this research, including advancements in technology, the emergence of other social media platforms, and shifts in the popularity of outdoor recreation-based lifestyles, also may play a role in this growth. The widespread adoption of smartphones equipped with sophisticated GPS systems has enhanced both security and accessibility to outdoor areas. This empowerment has encouraged outdoor enthusiasts to venture further into the backcountry and explore new destinations with increased confidence. Moreover, our findings linking increases in visits at some parks to online engagement, suggest that other social media platforms utilizing similar algorithmic-based content, such as YouTube and TikTok, may also exert influence on visitation as well. Additionally, the growing popularity of diverse outdoor activities such as rock climbing and thru-hiking (e.g., Appalachian Trail, Pacific Crest Trail), may have also shaped visitation patterns within the areas we studied. These evolving recreational preferences have contributed to the shifts in visitor numbers, highlighting the complex interplay of social media, technology, and changing outdoor lifestyles influencing park visitation trends. In other words, our evidence suggests some impact from Instagram that varies by park unit, but the app itself is likely not solely responsible for the record-breaking increased visitation to USNPs in the last decade as some traditional media outlets have claimed[4–7].

The findings of our study carry significant policy implications for both public land managers and social media companies. Social media platforms serve as cost-effective and potent tools for managers to actively engage with their visitors and disseminate information about park conditions, locations, policies, and practices. Our research highlights the other side of social media, demonstrating the influence wielded by influencer accounts producing content that receives high levels of user engagement. This behavioral channel through the bandwagon effect is likely impacting visitor patterns in specific areas. Our results suggest that park managers could monitor social media posts geotagged to their location and be more prepared for visitation increases if content associated with their location has a viral moment or their park unit is frequently geotagged by influencers. The NPS is actively trying to rein in certain behaviors linked to avid social media users that often lead to resource degradation. Currently, the NPS requires anyone with a monetized social media account to secure a commercial photography permit if the user is filming within the borders of over 400 NPS locations. The permits require self-identification from visitors who generate any income from content shared online. This policy is intended to help land managers better protect fragile resources, such as alpine meadows, from degradation by those seeking engaging content. However, enforcement of this policy is challenging for units with existing staffing shortages and many social media influencers are either unaware of the policy or unconcerned about any potential penalties.

Social media companies operate with relatively little regulation, yet the influence their apps and user bases have on the real world is profound. Integrating pertinent information about location-specific requirements or initiatives such as “Take only photographs, leave only footprints” directly within social media apps[52] could heighten user awareness about practices to minimize their impacts. Establishing a mechanism through which public land managers can request information about monetized accounts currently active on public lands could significantly enhance compliance enforcement with new NPS permitting requirements. Improved access to data from in-app indicators, such as viral content flags, associated with public lands could improve our understanding of this new evolving outdoor recreation landscape. By providing this pathway for information, social media companies could facilitate a deeper comprehension of the dynamics shaping contemporary outdoor activities, thereby empowering land managers to adapt and respond effectively to the shifting patterns of public land usage.

Methods

Visitation Data

USNP unit-specific monthly visitation data are publicly available as *Visitor Use Statistics* from NPS Stats.[53] The database contains information on visitor trips to over 400 park locations managed by the NPS. The type of visitor trip can include recreational or non-recreational visits, the visit hours and whether the trip included overnight stays. Our empirical setting focused on the impacts to visitation to national parks (NPs), which includes sixty-three (63) locations (Extended Data Fig. 1). For parsimony, we limit our analysis to NPs within the lower forty-eight states of the U.S. that are easily assessable by car (i.e., excluding island and water-based NPs). This reduces the sample to forty-five (45) NPs. The sample is further reduced to forty (40) parks since five parks were designated as a NP after 2010 (post-Instagram). In our final sample of forty NPs, there were six parks designated prior to 2010 but during our sample period (1993 - 2019; Great Sands Dunes, Cuyahoga Valley, Black Canyon of the Gunnison, Death Valley, Joshua Tree, and Saguaro NPs). For these parks, a binary variable indicating NP designation is included within our model to control for this change in status. Data on national park entrance fees from 1993 to 2019 were obtained through email communication with NPS. Our visitation dataset is a monthly panel covering a nearly 27-year period for 40 NPs.

Instagram Data

Social media data was collected by web scrapping all publicly available content geotagged to any USNP unit on Instagram from late 2019 back to the launch of the app in 2010. The content and metadata were collected using multiple Python scripts that were publicly available at the time of data collection. Readers interested in additional details about this data collection process are referred to Lowe Mackenzie et al. (2024)[21]. This process resulted in the collection of over 8.3 million unique observations. Each collected post includes metadata containing each NP's name and specific location. Any individual post falling outside the latitude and longitude coordinates of the park were identified and removed from the sample. The collected metadata also provided information on each post indexed by identification number, the date the photo was uploaded, the number of likes and comments the post received, and the caption entered by the poster. To ensure user privacy, we did not collect usernames, any information in any user's bio, or the specific image.

We transformed this data into metrics that can be used to represent different ways Instagram may influence visitation to USNPs. First, photo-user-day (PUD)[28] is metric that represents a unique upload from a user account observed at a park on a given day. The total PUD per month measures the contemporaneous volume of uploaded content to Instagram for each park. Second, engagement was quantified by the number of *likes* a photo receives. We measured the cumulative effect of Instagram posts with high engagement using the following equation:

$$Infl_{pT} = \sum_{t=1}^{T} Content_{pt}^{l_y} \ , \qquad \text{(M.1)}$$

where $Infl$ is the cumulative number of influential posts for park p at current time *(T)*, and $Content_{pt}^{l}$ is the content for park p in month *t* that meets threshold l by year, y. An influential post was quantified by surpassing a percentile threshold, l, based on the number of likes a post receives. Specifically, content was considered influential if it was in the top 90th or 95th percentile based on the total number of likes received on the entire set of photos geotagged per year (Extended Data Table 3).

Textual and Sentiment Analysis of Captions

We leveraged the depth and breadth of our Instagram data by cataloging and quantifying caption and hashtag information included in each geotagged post. Textual analysis provides an opportunity to turn text into data by systematically extracting meaning from fields of text[40,41]. Each of our approximately 8.3 million observations include a caption, which can contain a string of up to 2,200 characters in length. The text within these captions contains a range of information about the location, the experience, on-site activities or sentiment about the experience.

The process of textual analysis begins with cleaning and simplifying the text, employing techniques such as tokenizing, stemming, or lemmatizing. These methods are used to standardize the text, converting all letters to lowercase and eliminating stop words and punctuation marks from the dataset[40]. Subsequently, the focus shifts to extracting meaningful insights from the refined list of words. To accomplish this, we created a tailored dictionary for each park and assessed the sentiment of each post through natural language processing techniques. The park-specific dictionaries are curated with emphasis on three key sets of information: 1) geographical location, 2) specific points of interest within the park, and 3) activities associated with the park. For 1), each caption is processed to determine if the user is discussing the park or the state in which the park is located. A word frequency analysis to generate word clouds for each park location revealed this is a common practice as the park's name often emerges as the most prominent term (Extended Data Fig. 2). The second set of park-specific dictionaries leverages the National Park Service's Points of Interest (POI), a dataset that offers a comprehensive list of specific location names within each park categorized by types such as trailheads, viewpoints, historical buildings, restrooms, or parking lots. As an example, our POI dictionary includes specific landmarks like "Angels Landing" within Zion National Park, which holds a significant presence in Zion's word cloud (Extended Data Fig. 2). The third set of dictionaries focused on the diverse activities available within each park. We used each park's "Things-to-Do" section on their respective nps.gov websites to build a list of the activities listed. Activities across the 40 parks are quite varied and can include auto touring, backpacking, birdwatching, canyoneering, hiking, hot springs, photography opportunities, stargazing locations, swimming spots, and wildlife watching, among many others.

Next, we gauged the emotional tone and perspective (i.e., subjective, objective) of each post, providing insights into the sentiments expressed by users and the type of viewpoints expressed. We used the open-source Python library package Natural Language Toolkit (NLTK), and the modules Valence Aware Dictionary for sEntiment Reasoning (VADER) and Textblob[62]. VADER sentiment analysis tool is specifically designed for analyzing text data from social media and blends sentiment lexicon approaches with grammatical rules for expressing polarity and intensity. The sentiment intensity scoring uses a scale and normalization process to rate words such as “horrible” and “great” as well as text that frequently appears in online conversations, such as colloquialisms (lowkey), intense punctuation (!!), all caps emphasis (this is VERY awesome), emoticons (😍), as well as acronyms commonly used in online slang such as LOL (i.e., “laughing out loud”, used to express amusement). The process estimates the intensity of sentiment to determine whether the statement has an overall positive, negative, or neutral tone. Extended Data Table 1 provides a sample of the sentiment analysis results.

TextBlob is based on a pre-trained sentiment analysis model that uses a combination of a Naïve Bayes classifier and pattern-based approaches to classify the subjectivity of text. The library includes a score that ranges from 0 to 1, where 0 indicates the content is highly objective with factual information and 1 indicates highly subjective content classified as opinions. In other words, this subjectivity score provided a measure of how much of the text expresses opinions, emotions, or other subjective information as opposed to being purely factual or objective.

Count Data Model of Engagement

Our park-specific data dictionaries and sentiment and subjectivity analyses provide additional variables to specify a model estimating the factors that may impact engagement with posts on Instagram. We measured engagement as a non-negative count of the number of likes the content uploaded to Instagram recieved by the time the information was collected. The number of likes does not fit a normal distribution and exhibits a highly right-skewed pattern, suggesting a large amount of content recieves little to no engagement. The variance in the data is 52,000 times larger than the mean, indicating a significant overdispersion in the distribution. We selected a negative binomial regression for this analysis as the approach is well-suited for count data exhibiting overdispersion[63]. A Poisson regression was also estimated as a robustness check and yielded quantitatively similar results (Extended Data Fig. 6).

We specified the model estimating the impact of many factors on the number of likes ($Likes_{ipt}$) an individual (*i*) Instagram post at park p in month *t* received as follows:

$$Likes_{ipt} = \beta_1 Location_{ipt} + \beta_2 POI_{ipt} + \beta_3 Things2Do_{ipt} + \beta_4 Sentiment_{ipt} +$$

$$\beta_5 Subjective_{ipt} + \beta_6 Hashtags_{ipt} + \beta_7 Hashtags_{ipt}^2 + \beta_8 Num.Words_{ipt}$$

$$\beta_9 UserInt_{ipt}^{Q} + \rho_p * \tau_{y(t)} + \gamma_t + \varepsilon_{ipt}\, , \qquad \text{(M.2)}$$

where $Location_{ipt}$, POI_{ipt}, $Things2DO_{ipt}$ are binary indicators that are equal to one if the post contains text included in the three park-specific dictionaries, respectively. $Sentiment_{ipt}$ is a categorical variable using a determination from VADER on whether the caption expressed sentiment that was considered positive, negative or neutral. $Subjective_{ipt}$ is a value between 0 and 1 determined by Textblob capturing the extent of fact or opinion within the text. $Hashtags_{ipt}$ reprsent the number of hashtags used in the caption. In 2021, Instagram reported the optimal number of hashtags for content creators is 3-5 hashtags per post indiciating a non-linear relationship with the number of hashtags used in a individual post. We capture this potential effect by including a quadratic representation of the number of hastags included in the caption. $UserInt_{ipt}^{Q}$ is a binary variable controlling for the intensity level of posting for each user i at park p. Since we did not collect any personally identifiable information (i.e., we do not directly observe influencers or monetized accounts), we captured this variable as posting frequency at USNP geotags. On average, an individual user posts at least once with a NP geotag, while the highest observed count by a single user reaches 4,333. Due to computational constraints associated with the substantial number of observed individuals posting in our dataset (2.8 million), we use an approach to control for activity intensity within four quantile interval thresholds (Q). The low activity users post frequency falls between the 25th to 50th percentiles, while moderately active users post frequency falls between the 50th to 75th percentiles. High activity users post frequency falls between the 75th to 90th percentiles. Avid activity users are classified if their posting behavior was in the top 90th percentile. We explored user intensity classifciations further at the 95th and 99th percentiles as well (Extened Data Fig. 6). Lastly, $\rho_p * \tau_{y(t)}$ and γ_t represent park-by-month and year (y) fixed effects to control for seasonal changes in visitation at each park and time-invariant unobserved factors specific to each year, respectively.

Unsurprisingly, due to the large number of observations in our dataset, the negative binomial regression finds many statistically significant variables impacting engagement. As a sensitivity analysis, we used a random forests approach using a technique called feature importance to evaluate the contribution of each variable in the model in a different framework. A random forest is a machine learning method that builds multiple decision trees during training and outputs the average prediction of the individual trees for a regression problem. Key advantages of random forests include the ability to handle high-dimensional data, capture importance of relationships, and provide robust predictions[64,65]**.** Our analysis used 20% of the data for testing and 80% of the data for training the model. At each node in each decision tree, the Gini impurity was calculated before and after a split based on a particular feature. The decrease in impurity from a split was computed as the weighted average of impurity across all trees in the forest. The feature importance was calculated as the total decrease in impurity achieved by splits over that feature, averaged over all trees in the forest. This quantified the contribution of each variable to the overall predictive performance of the model, based on the decrease in impurity achieved by using that feature for splitting nodes in the decision trees comprised in each forest[65]. The results of this feature importance supported results from our negative binomial model specification, demonstrating that posts by avid users were a strong indicator of the level of engagement a post recieves. For completeness, partial dependence plots (PDPs)[66] are provided for interpreting the predictions of the random forest model and demonstrate the robustness of our findings (Extended Data Fig. 3). The PDP for user intensity provides supporting evidence for defining avid users as those in the 90^{th} percentile. A large spike occurs around 40 posts, which is the threshold number of posts for a user to be classified as a avid user (Extended Data Table 4). The results provide evidence supporting user intensity as a primary driver of online engagement.

Identification of Park Groups by Social Media Activity

Under the hypothesis that different content metrics may have heterogeneous effects across USNP units, we developed a process to identify parks with differing levels of social media exposure based metrics deriving from Instagram's user base: 1) the most liked photo and 2) the sum of posts from influencers at each park. Fig. 4 visually suggested potential grouping strategies based on these metrics. The y-axis position represents the number of likes for the highest-liked content for each park and the x-axis represents the sum of posts from influencers at each park. We use clustering methods to more formally identify potential groupings.

First, we employ a DBSCAN (Density-Based Spatial Clustering of Applications with Noise), a non-parametric clustering algorithm that identifies clusters based on local point density. This process helps identify potential outliers within our data of importance to viral impact. To ensure comparability across metrics, indicators (maximum likes and total influencer presence) are normalized to ensure equal weighting in analysis. DBSCAN identifies two clear outliers, Yosemite and Zion National Parks (Extended Data Fig. 7.a). To understand the underlying cluster structure of the remaining parks, a k-medians approach is then employed using a Manhattan distance, which is more robust to outliers compared to Euclidean distance. Since k-medians require a priori specification of the number of clusters, the silhouette method is used to determine the optimal number. Results from the silhouette criterion suggest two additional clusters. Together, these results provide support for our three-group clustering strategy. The high exposure group includes the significant outliers: Yosemite and Zion NPs. The moderate exposure group contained thirteen (13) parks: Arches, Bryce Canyon, Death Valley, Glacier, Grand Canyon, Grand Teton, Great Smoky Mountains, Joshua Tree, Kings Canyon, Mount Rainier, Olympic, Rocky Mountain, and Sequoia. The remaining twenty-five (25) parks were designated as the low exposure group with their most viral post under 1 million likes and with less than 30,000 posts from influencers. Extended Data Figs. 4 and 5 show the most viral content and number of influencers at each park, respectively.

Additional Data for Visitation Models

The monthly panel structure of our data necessitated the inclusion of additional covariates at the same time step for our visitation models. Macroeconomic variables that may influence visitation trends were retrieved from the Federal Reserve Bank of St. Louis[54]. We collected a seasonally adjusted national monthly income level measuring real disposable personal income (RDPI)[55] and the state-level unemployment rates, which captured a measure of consumer confidence[27,54]. State population totals are available as an annual estimate[54,56] and we use a linear interpolation to generate monthly variation to match the time step of our model. For each variable provided at a state level, the measure used is connected to the park units within those state boundaries. To control for the changing demographics of the general population, the percent of the population 65 and above was obtained from the U.S. Census Bureau.

Gasoline prices can impact individual travel costs and have been found to be significant component in visitation models[25,27]. NPs are often accessed by car and gasoline prices may reflect budget constraints on those choosing whether to take a trip to a NP. Historical monthly average real gasoline prices adjusted for inflation were obtained from the U.S. Energy Information Administration (EIA)[57]. Lastly, weather conditions may also affect visits to recreation sites[58–60]. Daily observations of maximum temperature and precipitation for all parks were collected from PRISM Climate Group at Oregon State University[61]. Exploiting this daily variation, weather enters our model as a non-linear response function using a binning approach[58,59], where we sum the number of days per month in each bin. The temperature bins are set at a 10-degree Fahrenheit scale (i.e., 30.0 – 39.99°F, 40.0 – 49.99°F, etc.) and the precipitation bins represent days with observed levels of perception in inches (i.e. no rain 0”, low 0” >.5”, moderate .5”>1.5”, high 1.5”>). We use maximum temperature instead of daily mean temperatures for behavioral and spatial reasons.

Information broadcasted on weather stations provided to potential recreators is reported as a daily high and low for a region. Therefore, recreators are likely making decisions based on maximum daily temperature rather than the mean. Another spatial concern arises given the national scope of this analysis. Locational variance in temperature is heterogenous across regions. Deserts areas can have very hot days as well as very cold nights, and such extremes could dilute an average measure. In such a case, a desert and a forested area with milder weather would have a similar mean temperature distribution, resulting in some measurement error. Using daily maximum temperature alleviates this issue.

Visitation Models

Using our USNP groupings based on viral content and the presence of influencers posting at park units, we estimate a set of visitation models to determine the impacts of Instagram holistically and then using elements of engagement and types of users to test our two hypotheses. The baseline model specification examined first is to test the general impact of Instagram's rise to popularity on visitation at park p in month t ($Visit_{pt}$) using a binary indicator to denote the post-Instagram period, IG_t. The baseline model was specified as follows:

$$Ln\left(Visit_{pt}\right) = \beta_0 + \beta_1 IG_t + \beta_2 IG_t \; x \; SMG\_H_p + \beta_3 \mathbf{T}_{pt} + \beta_4 \mathbf{PRE}_{pt} + \beta_5 \mathbf{X}_t$$

$$\beta_6 Ln(Fee)_{pt} + \beta_7 Desig_{pt} + \rho_p * \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \; . \qquad \text{(M.3)}$$

In eq. M.3, visits per park per month are log-transformed to reduce the impact of outliers. β_2 is our coefficient of interest as the effect on visitation after ther rise of Instagram at parks with high social media activity (SMG_H_p), initially as defined by Wichman (2024) for comparison. $\mathbf{T}_{pt}$ represents daily counts of days per month in observed maximum temperature bins at each park and $\mathbf{PRE}_{pt}$ represents the number of days each month in one of four (4) precipitation bins. Additional controls include $\mathbf{X}_t$, a vector of non-park specific covariates (e.g., national income, unemployment rate, etc) in month t, $Ln(Fee)_{pt}$, the natural log of park-specific entrance fees, and $Desig_{pt}$, a binary indicator for NP designation if the park was designated as a National Park during the sample frame. The panel nature of our data allowed for additional controls for unobserved time-invariant impacts using fixed effects. As such, $\rho_p * \tau_{y(t)}$ and γ_t control for park- by-month and year fixed effects, respectively.

Similar to Wichman (2024)[22], we adopt an instrumental variables (IV) approach to overcome endogeneity concerns due to potential correlation between visitation and social media activity. Using Google Trends to develop a measure of online popularity at each USNPs before social media, we can account for the relationship and potentially remove the endogeneity concerns. The instrument created here is correlated with social media exposure but should not directly impact visitation after the release of Instagram, allowing for plausibly causal estimates of the impact of social media activity on USNP visitation. To first compare our results to previous research, we replicate the decisions of Wichman (2024) in exposure groups and model specification and find their result of a large significant increase in visitation at high social media exposure parks is sensitive to the choice of clustering on the standard errors. A two-way clustering approach by park-month assumes observations from the same park but in different months are independent. It also assumes clustering at both the temporal and spatial units[43,45,46]. Implementing a multi-way cluster in this fashion would also mean the clustered errors intersection are partially nested dimensions and would not be independent of each other. Given the treatment of influence from social media is highly correlated with parks overtime, failing to account for this within-park serial correlation would lead to over rejection of the null[45]. The clustering level suggested in the literature would be solely determined by the sampling process and treatment assignment[44]. Extended Data Table 2 highlights the two-way clustering on park-month and year-month produces standard errors comparable in magnitude to heteroskedasticity robust estimates when using the IV strategy. When clustering on park and year-month, the standard errors double and lose statistical significance, suggesting nontrivial within-park correlation overtime. This sensitivity analysis suggests the need for a conservative multi-cluster on non-nested dimensions of the spatial unit (park) to capture correlations within each park across time and year-month to capture correlation across parks within each year month period. This conservative approach leads to inconclusive results for impacts to visitation from the rise of Instagram generally using an IV strategy.

To test further explore the potential connections between Instagram and visitation and to test our hypotheses, reduction in search and information costs and herd behavior, as potential reasons why social media exposure may influence visitation, we refine the visitation model in multiple dimensions. First, we apply our three social media exposure groups and add both PUDs[28] to capture a contemporaneous effect and the sum of all prior Instagram posts with high engagement (eq. M.1) to capture a cumulative effect of viral content. We used IG_{Launch_t} to distinguish between months in which a park received zero geotagged Instagram posts (2010 – 2019) and months during which Instagram had not yet launched (1993 – 2010). The subscript G to the *PUD* and *Infl* terms reflect the high, moderate, and low exposure groups determined in Fig. 4. We use this strategy to isolate parks with geotagged viral engagement online to see if there are differential impacts across park types for both content variables. We specify our preferred model as:

$$Ln(Visit_{pt}) = \beta_0 + \beta_1 IG_{Launch_t} + \beta_2 PUD_{pGt} + \beta_3 Infl_{pGT}{}^{j} + \beta_4 \mathbf{T}_{pt} + \beta_5 \mathbf{PRE}_{pt} + \beta_6 \mathbf{X}_t + \beta_7 Ln(Fee)_{pt} + \beta_8 Desig_{pt} + \rho_p * \tau_{y(t)} + \gamma_t + \varepsilon_{pt} . \quad (4)$$

The superscript *j* on the *Infl* term indicates that this variable can be used to estimates cumulative impacts of influential content (i.e., *j*=1) or influential users (i.e., *j*=2). We isolated viral or influential content from avid users to determine if the impacts of such content were driven by a particular set of Instagram users. Avid users account for 27% of the influential content determined by eq. M.2 with the 90th percentile threshold. We compared results from our original subset of the 90^{th} percentile of engagement (Fig. 3.d) to results when differentiating engaging content coming from avid and non-avid users. This comparative exercise demonstrated that Instagram posts that go viral online can systematically impact visitation at parks like Yosemite and Zion regardless of user type, but that avid users with viral content have the potential to increase visitation at a broader set of parks than non-avid users.

1. Timmons, A. L. Too Much of a Good Thing: Overcrowding at America's National Parks. *Notre Dame Law Rev.* **94**,.
2. Congressional Research Service. *National Park Service Deferred Maintenance: Overview and Issues.* https://crsreports.congress.gov (2024).
3. Robbins, J. How A Surge in Visitors Is Overwhelming America's National Parks. *Yale E360* https://e360.yale.edu/features/greenlock-a-visitor-crush-is-overwhelming-americas-national-parks (2017).
4. Solomon, C. Is Instagram Ruining the Great Outdoors? *Outside Online* https://www.outsideonline.com/culture/opinion/instagram-ruining-great-outdoors/ (2017).
5. Simmonds, C., McGivney, A., Wilkinson, T. & Canon, G. Crisis in our national parks: how tourists are loving nature to death. *The Guardian* (2018).
6. Hegyi, N. Instagramming Crowds Pack National Parks : NPR. https://www.npr.org/2019/05/28/726658317/instagramming-crowds-pack-national-parks (2019).
7. Lake, Z. & Rulli, M. National parks officials grappling with high volume as Instagram tourism booms. *ABC News* https://abcnews.go.com/Lifestyle/instagram-tourism-booms-horseshoe-bend-national-parks-officials/story?id=64638198.
8. Allcott, H. & Gentzkow, M. Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **31**, 211–236 (2017).
9. Guess, A. M. *et al.* How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* **381**, 398–404 (2023).
10. Allcott, H. *et al.* The effects of Facebook and Instagram on the 2020 election: A deactivation experiment. *Proc. Natl. Acad. Sci.* **121**, e2321584121 (2024).

11. Keles, B., McCrae, N. & Grealish, A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *Int. J. Adolesc. Youth* **25**, 79–93 (2020).

12. McLoughlin, K. L. *et al.* Misinformation exploits outrage to spread online. *Science* **386**, 991–996 (2024).

13. Da Silva Pinho, A., Céspedes Izquierdo, V., Lindström, B. & Van Den Bos, W. Youths' sensitivity to social media feedback: A computational account. *Sci. Adv.* **10**, eadp8775 (2024).

14. Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**, 554–559 (2016).

15. Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S. & Lewandowsky, S. Psychological inoculation improves resilience against misinformation on social media. *Sci. Adv.* **8**, eabo6254 (2022).

16. Epstein, Z., Sirlin, N., Arechar, A., Pennycook, G. & Rand, D. The social media context interferes with truth discernment. *Sci. Adv.* **9**, eabo6169 (2023).

17. Cinelli, M. *et al.* The COVID-19 social media infodemic. *Sci. Rep.* **10**, 16598 (2020).

18. Allen, J., Watts, D. J. & Rand, D. G. Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science* **384**, eadk3451 (2024).

19. Pierri, F. *et al.* Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Sci. Rep.* **12**, 5966 (2022).

20. Balmford, A. *et al.* Walk on the Wild Side: Estimating the Global Magnitude of Visits to Protected Areas. *PLOS Biol.* **13**, e1002074 (2015).

21. Lowe Mackenzie, A. L., Dundas, S. J. & Zhao, B. The Instagram Effect: Is Social Media Influencing Visitation to Public Land? *Land Econ.* 122920-0192R1 (2023) doi:10.3368/le.100.2.122920-0192R1.

22. Wichman, C. J. Social media influences National Park visitation. *Proc. Natl. Acad. Sci.* **121**, e2310417121 (2024).

23. Clawson, M. Outdoor recreation: Twenty-five years of history, twenty-five years of projection. *Leis. Sci.* **7**, 73–99 (1985).

24. Weiss, T. Tourism in America before World War II. *J. Econ. Hist.* **64**, (2004).

25. What Does the Future Hold for U.S. National Park Visitation? Estimation and Assessment of Demand Determinants and New Projections. *J. Agric. Resour. Econ.* (2020) doi:10.22004/ag.econ.298433.

26. Stevens, T. H., More, T. A. & Markowski-Lindsay, M. Declining National Park Visitation: An Economic Analysis. *J. Leis. Res.* **46**, 153–164 (2014).

27. Poudyal, N. C., Paudel, B. & Tarrant, M. A. A time series analysis of the impact of recession on national park visitation in the United States. *Tour. Manag.* **35**, 181–189 (2013).

28. Wood, S. A., Guerry, A. D., Silver, J. M. & Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **3**, 2976 (2013).

29. Sessions, C., Wood, S. A., Rabotyagov, S. & Fisher, D. M. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *J. Environ. Manage.* **183**, 703–711 (2016).

30. Fisher, D. M. *et al.* Recreational use in dispersed public lands measured using social media data and on-site counts. *J. Environ. Manage.* **222**, 465–474 (2018).

31. Wilkins, E. J., Wood, S. A. & Smith, J. W. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environ. Manage.* **67**, 120–132 (2021).
32. Wood, S. A. *et al.* Next-generation visitation models using social media to estimate recreation on public lands. *Sci. Rep.* **10**, 15419 (2020).
33. Hausmann, A. *et al.* Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conserv. Lett.* **11**, e12343 (2018).
34. Walden-Schreiner, C., Leung, Y.-F. & Tateosian, L. Digital footprints: Incorporating crowdsourced geographic information for protected area management. *Appl. Geogr.* **90**, 44–54 (2018).
35. Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C. & Leung, Y.-F. Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio* **47**, 781–793 (2018).
36. Ghermandi, A. & Sinclair, M. Passive crowdsourcing of social media in environmental research: A systematic map. *Glob. Environ. Change* **55**, 36–47 (2019).
37. Miller, Z. D., Taff, B. D., Newman, P. & Lawhon, B. A Proposed Research Agenda on Social Media's Role in Visitor Use and Experience in Parks and Protected Areas. *J. Park Recreat. Adm.* (2019) doi:10.18666/JPRA-2019-9553.
38. Stigler, G. J. The Economics of Information. *J. Polit. Econ.* **69**, 213–225 (1961).
39. Leibenstein, H. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *Q. J. Econ.* **64**, 183 (1950).
40. Gentzkow, M., Kelly, B. & Taddy, M. Text as Data. *J. Econ. Lit.* **57**, 535–574 (2019).

41. Dugoua, E., Dumas, M. & Noailly, J. Text as Data in Environmental Economics and Policy. *Rev. Environ. Econ. Policy* **16**, 346–356 (2022).

42. NPS.gov Homepage (U.S. National Park Service). https://www.nps.gov/index.htm.

43. Colin Cameron, A. & Miller, D. L. A Practitioner's Guide to Cluster-Robust Inference. *J. Hum. Resour.* **50**, 317–372 (2015).

44. Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. When Should You Adjust Standard Errors for Clustering? *Q. J. Econ.* **138**, 1–35 (2022).

45. Bertrand, M., Duflo, E. & Mullainathan, S. How Much Should We Trust Differences-In-Differences Estimates? *Q. J. Econ.* **119**, 249–275 (2004).

46. MacKinnon, J. G., Nielsen, M. Ø. & Webb, M. D. Cluster-robust inference: A guide to empirical practice. *J. Econom.* **232**, 272–299 (2023).

47. Weiler, S. & Seidl, A. What's in a Name? Extracting Econometric Drivers to Assess The Impact of National Park Designation*. *J. Reg. Sci.* **44**, 245–262 (2004).

48. Weiler, S. A park by any other name: National Park designation as a natural experiment in signaling. *J. Urban Econ.* **60**, 96–106 (2006).

49. Schwartz, Z. & Lin, L.-C. The impact of fees on visitation of national parks. *Tour. Manag.* **27**, 1386–1396 (2006).

50. Watson, C. An Economic Analysis of National Park Visitation Rates. (Montana State University, 2013).

51. Flores, D. & Kuhn, K. Latino Outdoors: Using Storytelling and Social Media to Increase Diversity on Public Lands. *J. Park Recreat. Adm.* **36**, (2018).

52. Krishnan, N., Gu, J., Tromble, R. & Abroms, L. C. Research note: Examining how various social media platforms have responded to COVID-19 misinformation. *Harv. Kennedy Sch. Misinformation Rev.* (2021) doi:10.37016/mr-2020-85.

53. STATS - Welcome to Visitor Use Statistics. https://irma.nps.gov/Stats/.

54. Federal Reserve Economic Data | FRED | St. Louis Fed. https://fred.stlouisfed.org/.

55. McIntosh, C. R. & Wilmot, N. An Empirical Study of the Influences of Recreational Park Visitation: The Case of US National Park Service Sites. *Tour. Econ.* **17**, 425–435 (2011).

56. Bureau, U. C. Data. *Census.gov* https://www.census.gov/data.

57. Homepage - U.S. Energy Information Administration (EIA). https://www.eia.gov/.

58. Dundas, S. J. & von Haefen, R. H. The importance of data structure and nonlinearities in estimating climate impacts on outdoor recreation. *Nat. Hazards* **107**, 2053–2075 (2021).

59. Dundas, S. J. & von Haefen, R. H. The Effects of Weather on Recreational Fishing Demand and Adaptation: Implications for a Changing Climate. *J. Assoc. Environ. Resour. Econ.* **7**, 209–242 (2020).

60. Chan, N. W. & Wichman, C. J. Climate Change and Recreation: Evidence from North American Cycling. *Environ. Resour. Econ.* **76**, 119–151 (2020).

61. PRISM Climate Group at Oregon State University. https://prism.oregonstate.edu/.

62. Hutto, C. & Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. Web Soc. Media* **8**, 216–225 (2014).

63. Wooldridge, J. Econometric Analysis of Cross Section and Panel Data. (2010).

64. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

65. Lior Rokach & Oded Z Maimon. *Data Mining With Decision Trees: Theory And Applications*. (World Scientific Publishing Company, Singapore, SINGAPORE, 2007).

66.Greenwell, B., M. pdp: An R Package for Constructing Partial Dependence Plots. *R J.* **9**, 421 (2017).

Data Availability: All data used in this study will be made publicly available in an open repository (e.g., Zenodo) before publication to comply with the journal's data availability policy. Data is currently available for the review process in a GitHub repository maintained by the corresponding author: https://github.com/loweas/nps

Code Availability: Replication code for this study, including instructions on how to obtain, process, and analyze Instagram data, will be made publicly available in an open repository (e.g., Zenodo) before publication to comply with the journal's data availability policy. Data is currently available for the review process in a GitHub repository maintained by the corresponding author: https://github.com/loweas/nps

Acknowledgements: No funding was received to support this research.

Author Contributions: ACLM: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing- Original draft preparation, Visualization.

SJD: Conceptualization, Methodology, Writing- Reviewing and Editing, Validation, Project Administration.

Competing Interests: The authors declare no competing interests

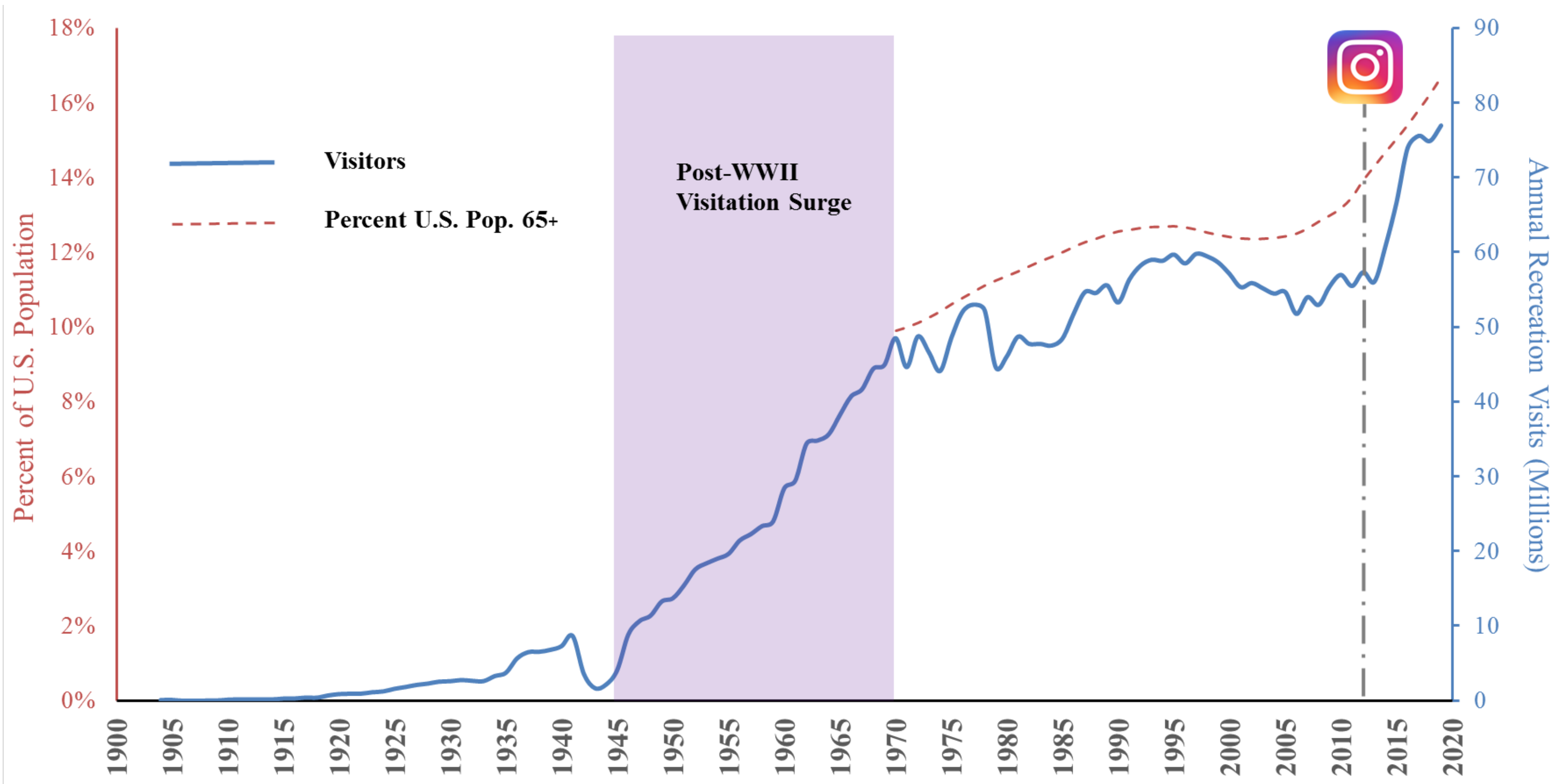


Fig. 1 | Annual recreational visits at U.S. National Parks 1904 to 2020. Right y-axis in blue is total annual recreation visitors in millions and the left y-axis in red is the percentage of the US population 65 or older.

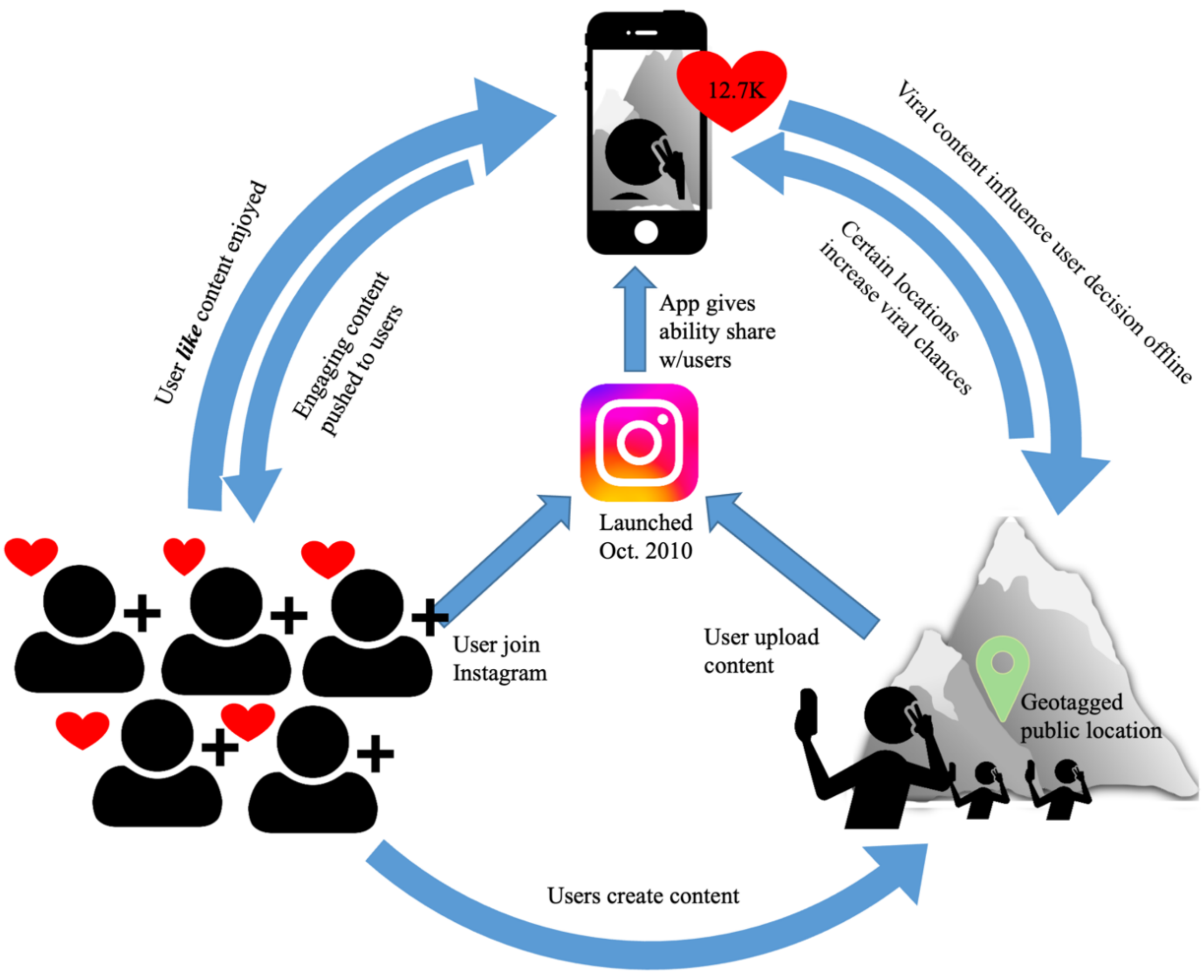


Fig. 2 | Conceptual diagram of Instagram's influence of visitation to public lands. When social media platforms begin, it takes time to develop a user base that uploads content and engages with other users' content. Instagram launched in October 2010 and reached 50 million users by April 2012. As content is uploaded to the platform, specific content engages the user base at higher levels than others. There are many potential motivations and outcomes associated with user engagement and the resulting influential content may affect some user decisions outside of the app. In context of visitation to public lands, viral content could influence those exposed to a new location to want to see it in person. Content with locational information, such as GPS coordinated (geotags), provides information on where to go to find a location or get a similar photo and may contain information about the quality of the location. Content posted by influencers, or frequent and popular app users, may also drive visitation through a behavioral channel (i.e., bandwagon effect).

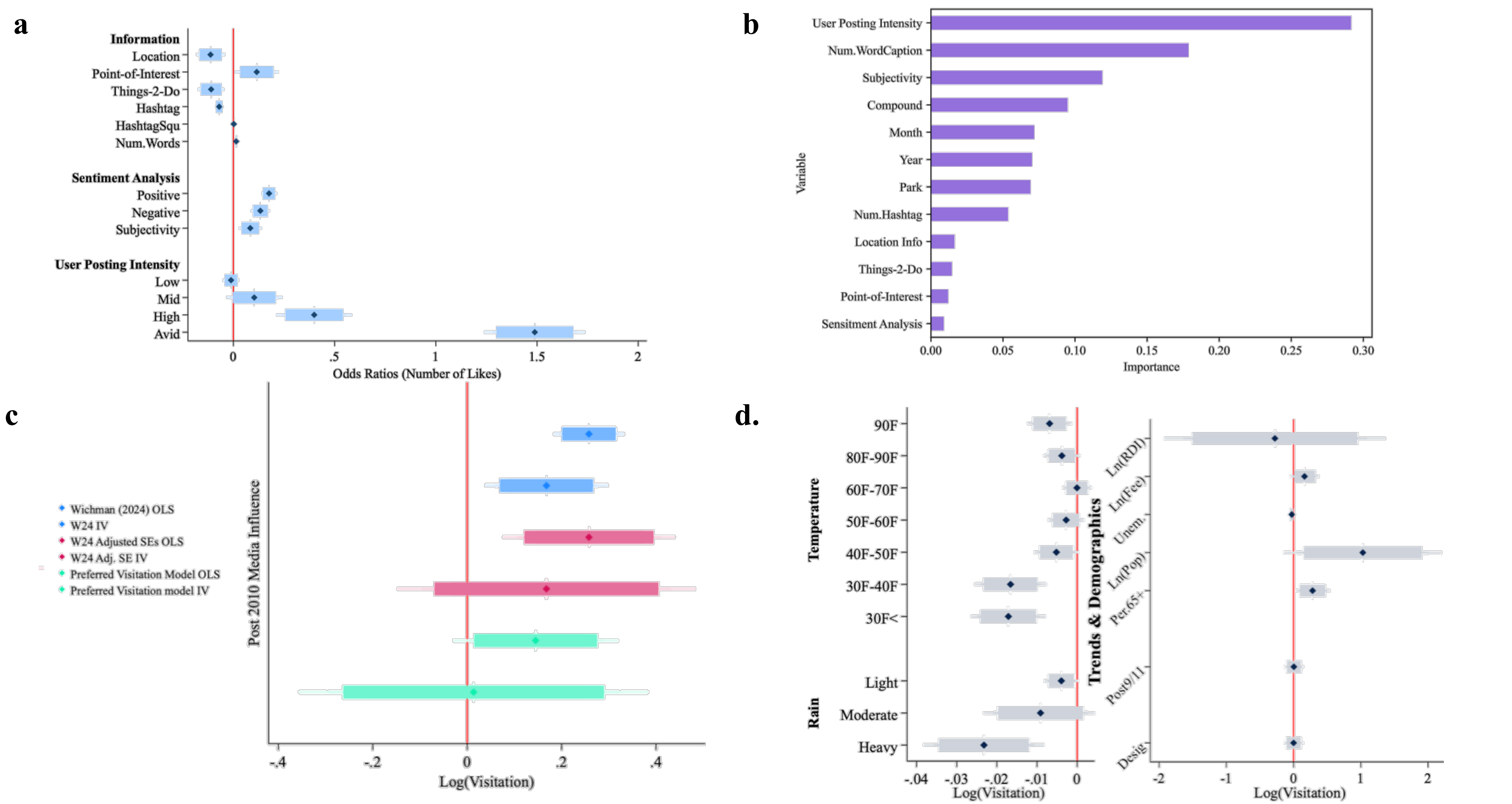


Fig. 3 | Factors influencing online engagement visitation trends. Colored rectangles indicate 95 % confidence intervals (CIs) of estimates and the extended lines show 99% CIs in panels (**a**), (**b**) and (**d**). Panel (**a**) Results suggest that high and avid user posting intensity are the largest driver of in-app engagement. Panel (**b**) The feature importance plot from a random forest model suggests user posting intensity has the highest importance rating in the engagement model. Panel (**c**) Results from broad visitation models, suggesting that we find no statistically significant effect of Instagram at high social media exposure parks in primary IV specification (green, bottom), counter to Wichman (2024) (blue, top). Panel (**d**) Visitation model results for all other control variables (Methods).

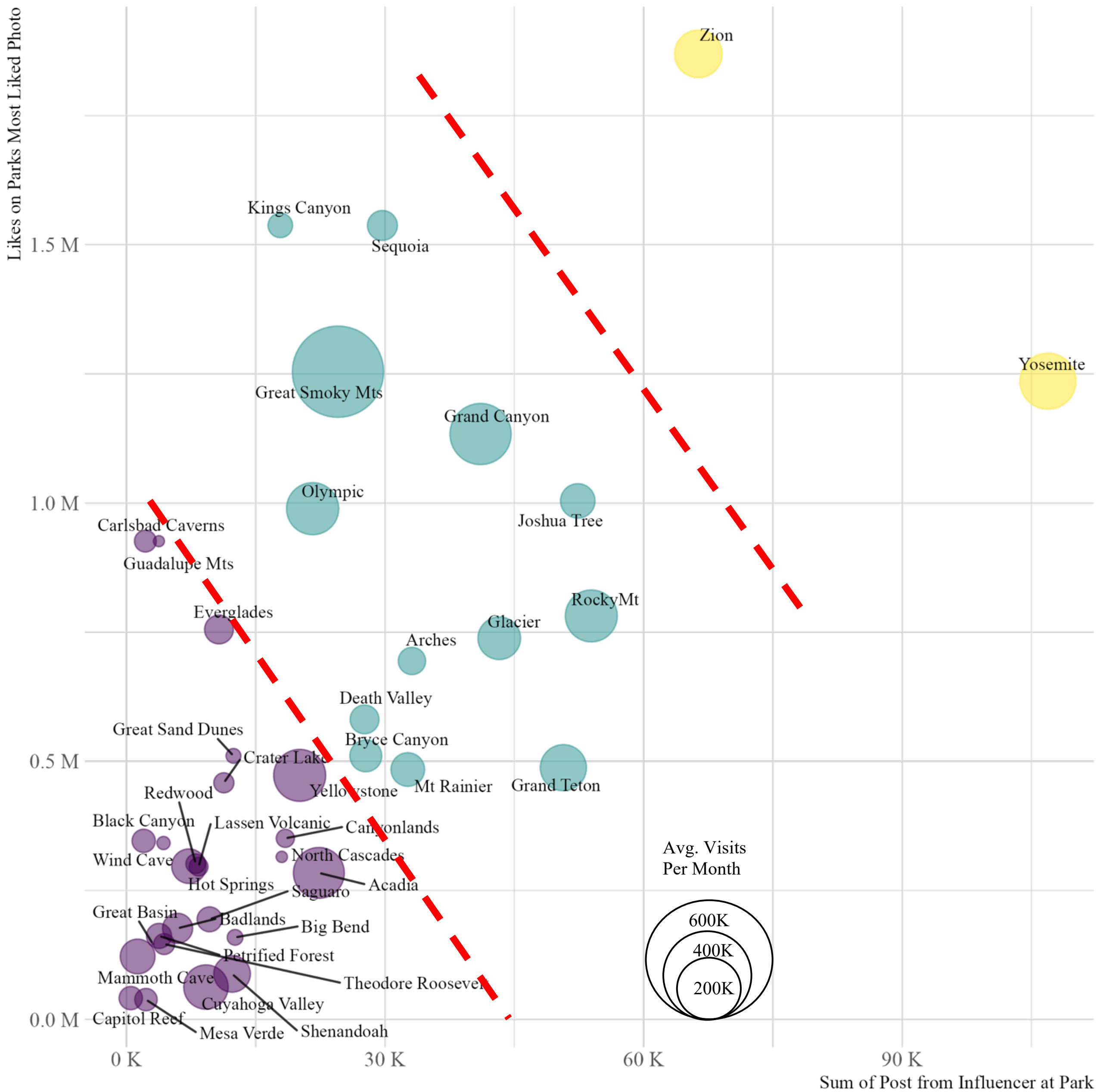


Fig. 4 | Modeling viral moments and visitation at US National Parks 2010 - 2019. Y-axis is the number of likes that the most liked photo from each park received. X-axis is the summation of the number of influencers observed at each park. The size of the circle represents average monthly visitation to park. Color identifies k-median cluster of groups based on two dimensions in the axes. Three grouping strategies are suggested in this plot. A High Engagement Group (yellow) had the most viral content and influencer presence on the app (Zion & Yosemite NPs). A Moderate Engagement Group (green) contains parks with less viral content and influencers than Zion or Yosemite, but still higher than many parks (Kings Canyon, Sequoia, Grand Canyon, Joshua Tree, Great Smoky Mountians, Olympic, Glacier, Death Valley, Grand Teton, Bryce Canyon, Mount Rainer, Rocky Mountain and Arches NPs). A Low Engagement group (purple) contains all other parks in our study area. Dotted red line represents the same thresholds represented in Extended Data Figs. 3-4.

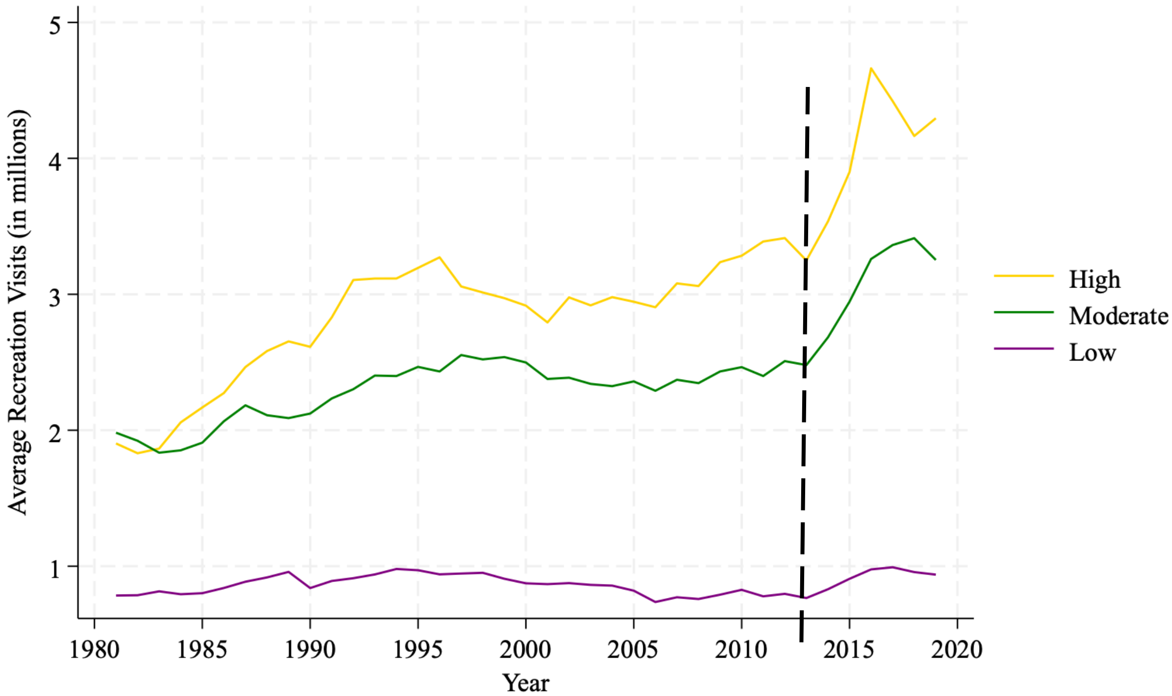


Fig. 5 | Visitation trends by social media engagement group. Y-axis is the average annual visits and the x-axis is time in annual increments. The vertical dotted line represents approximate timing of Instagram becoming a popular smartphone app (April 2012)[15]. The high engagement group (yellow) sees more average visitation than the other groups and saw a sharp increase post Instagram. The moderate engagement group (green) also shows modest growth after Instagram gained a wide user base. The trend in the low engagement group (purple) was realtively steady over 26 year timeline, including after Instagram's rise in popularity.

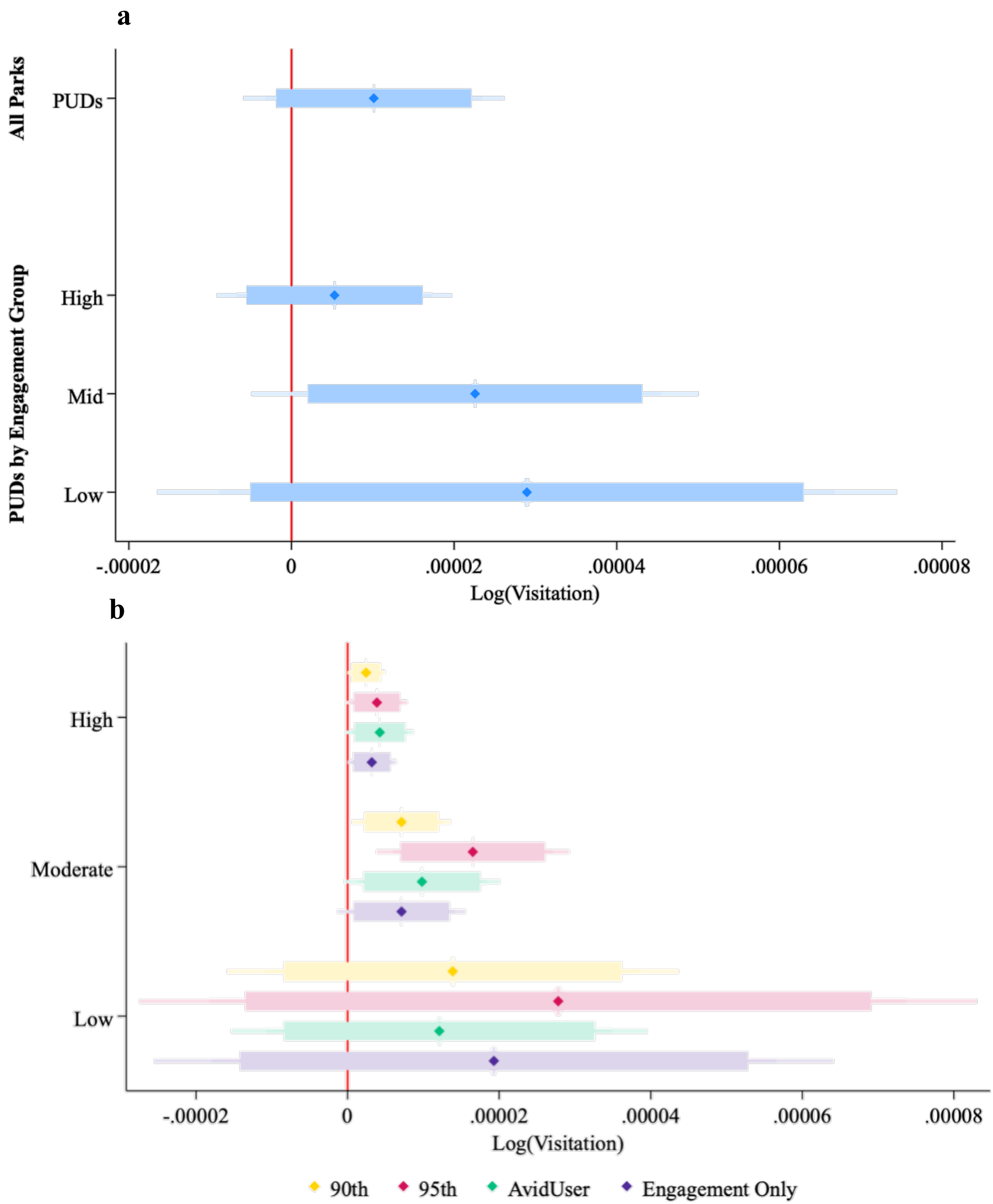
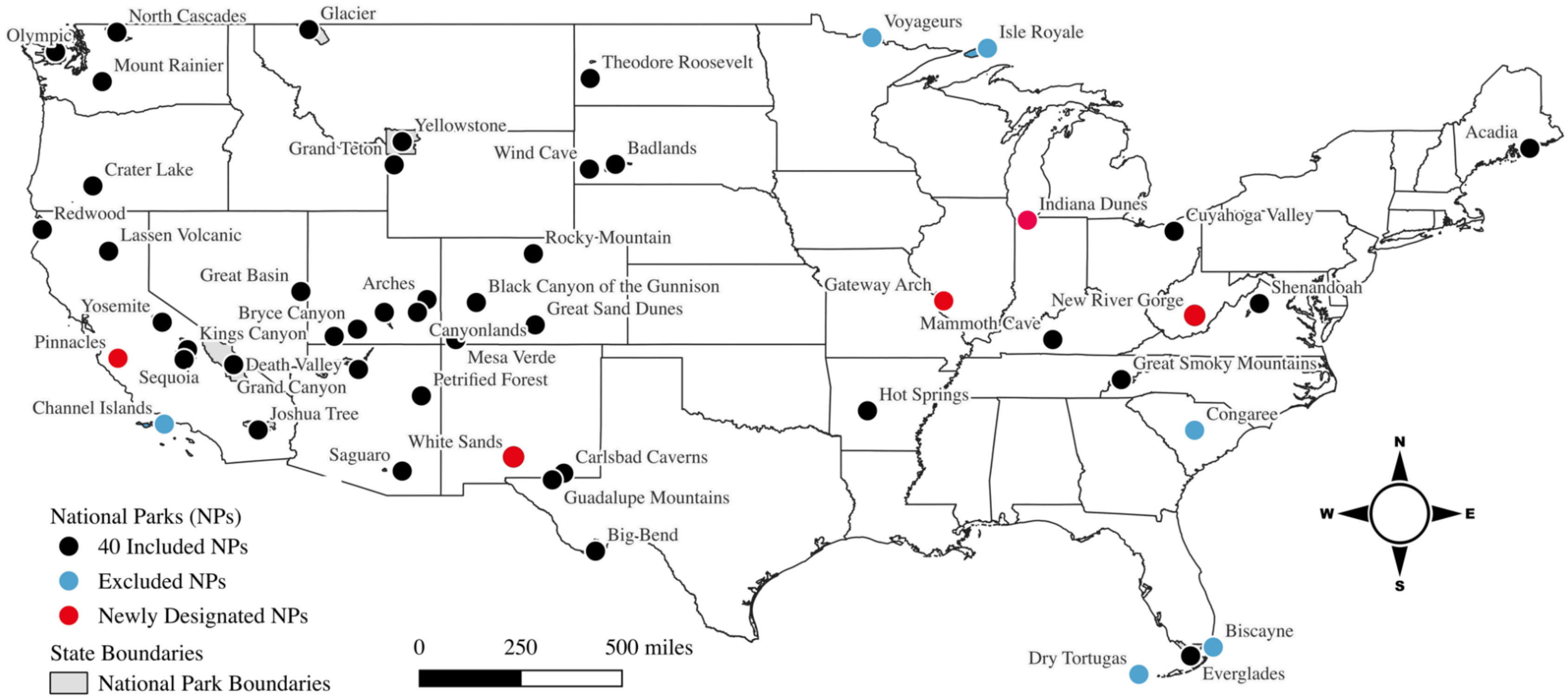
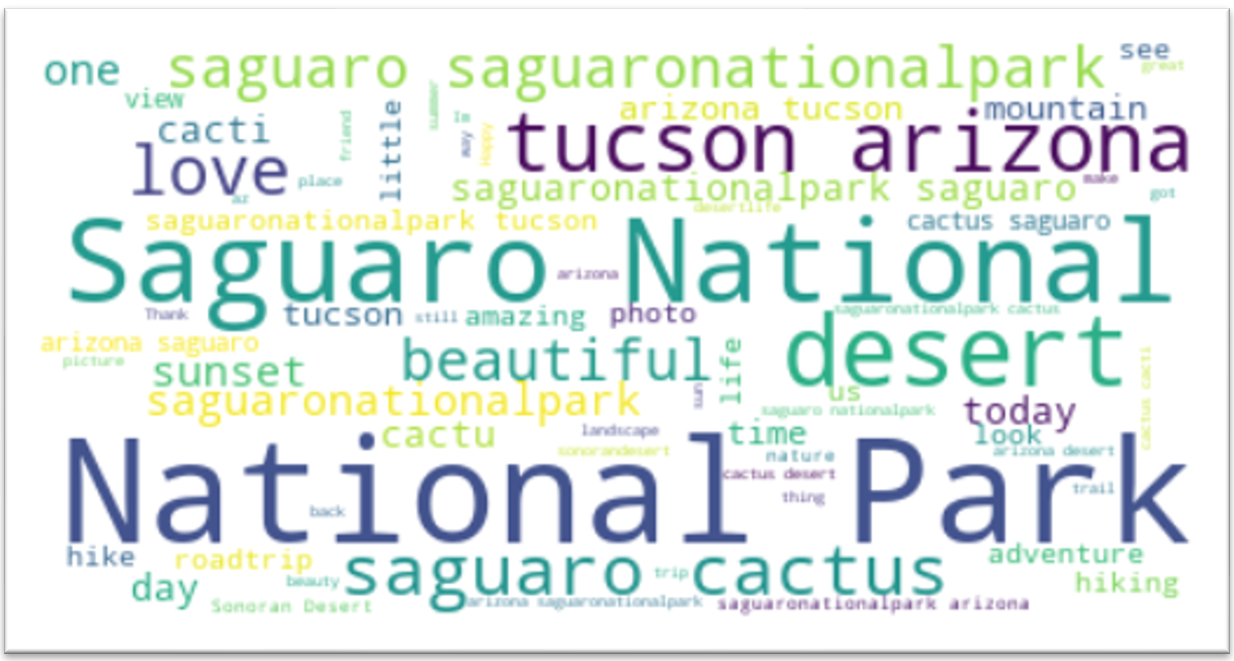


Fig. 6 | Factors determining engagement and the effect of avid users on visitation. Colored rectangles indicate 95 % confidence intervals (CIs) of estimates and the extended lines show 99% CIs. Panel (**a**): shows results of PUDs on visitation. Panel (**b**) reports results modeling the effect of posts in the 90th (95th) percentile of highly engaging content in yellow (red), posts from avid users in the highly engaging content category (green) and engaging posts from non-avid users (purple).

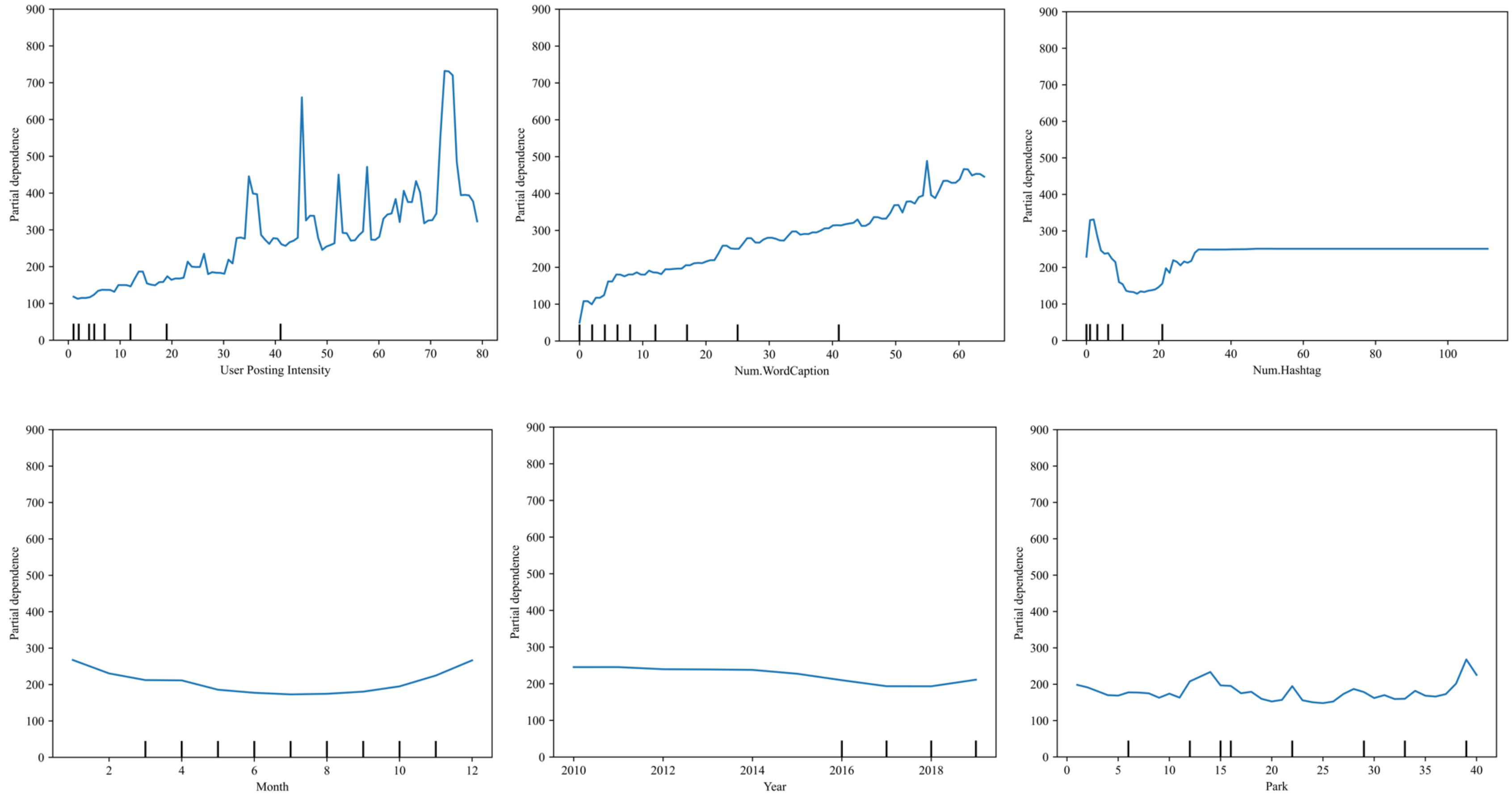
Extended Data Figures and Tables



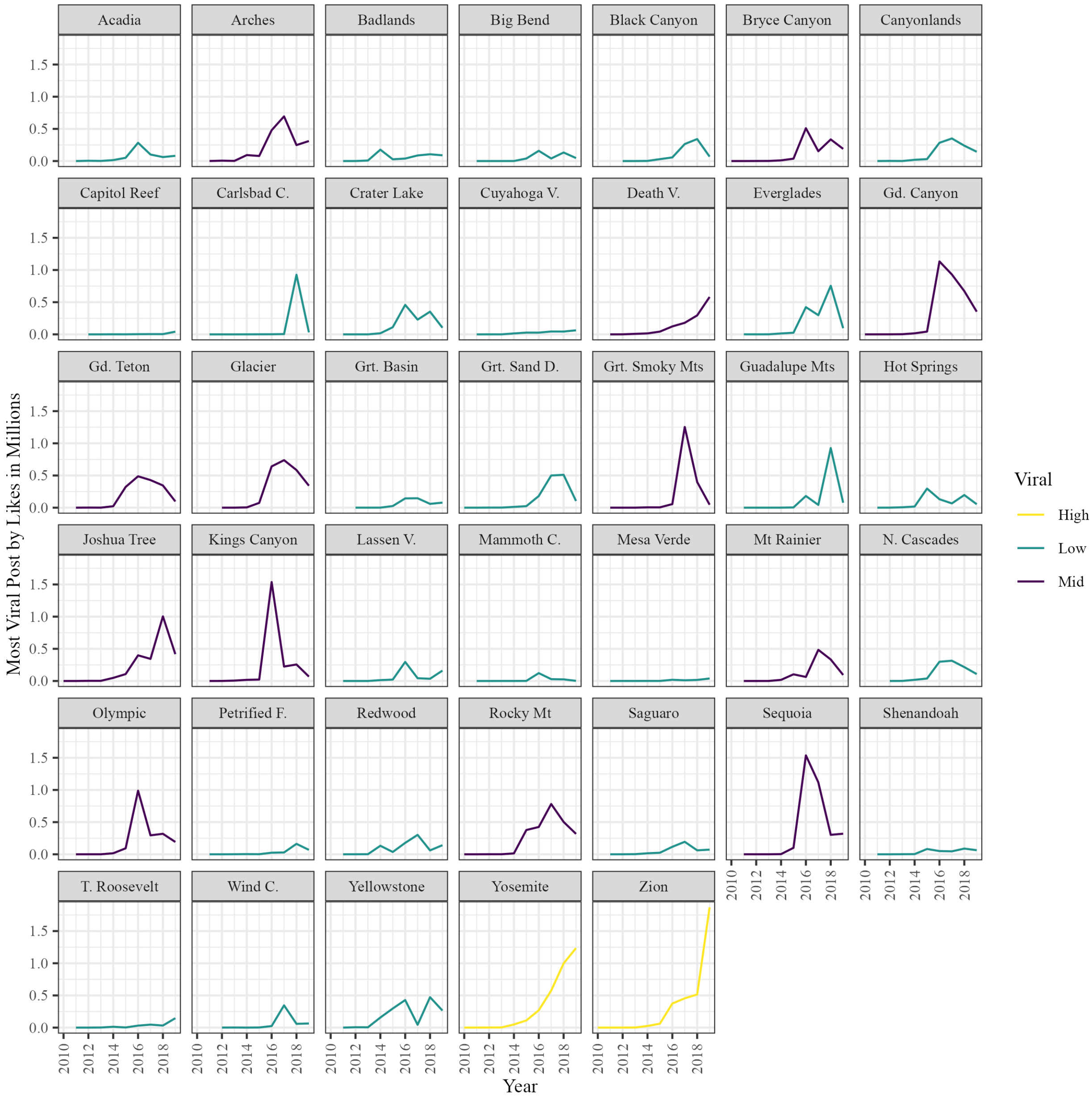
Extended Data Fig. 1: Map of National Parks in the continental United States. Map provides a visualization of the 40 NPs used in this paper (black circles). Red circles are NPs that were classified as NPs post-Instagram and blue circle NPs are parks with primarily water access, so they are not included in the study. National Park boundaries are gray areas.



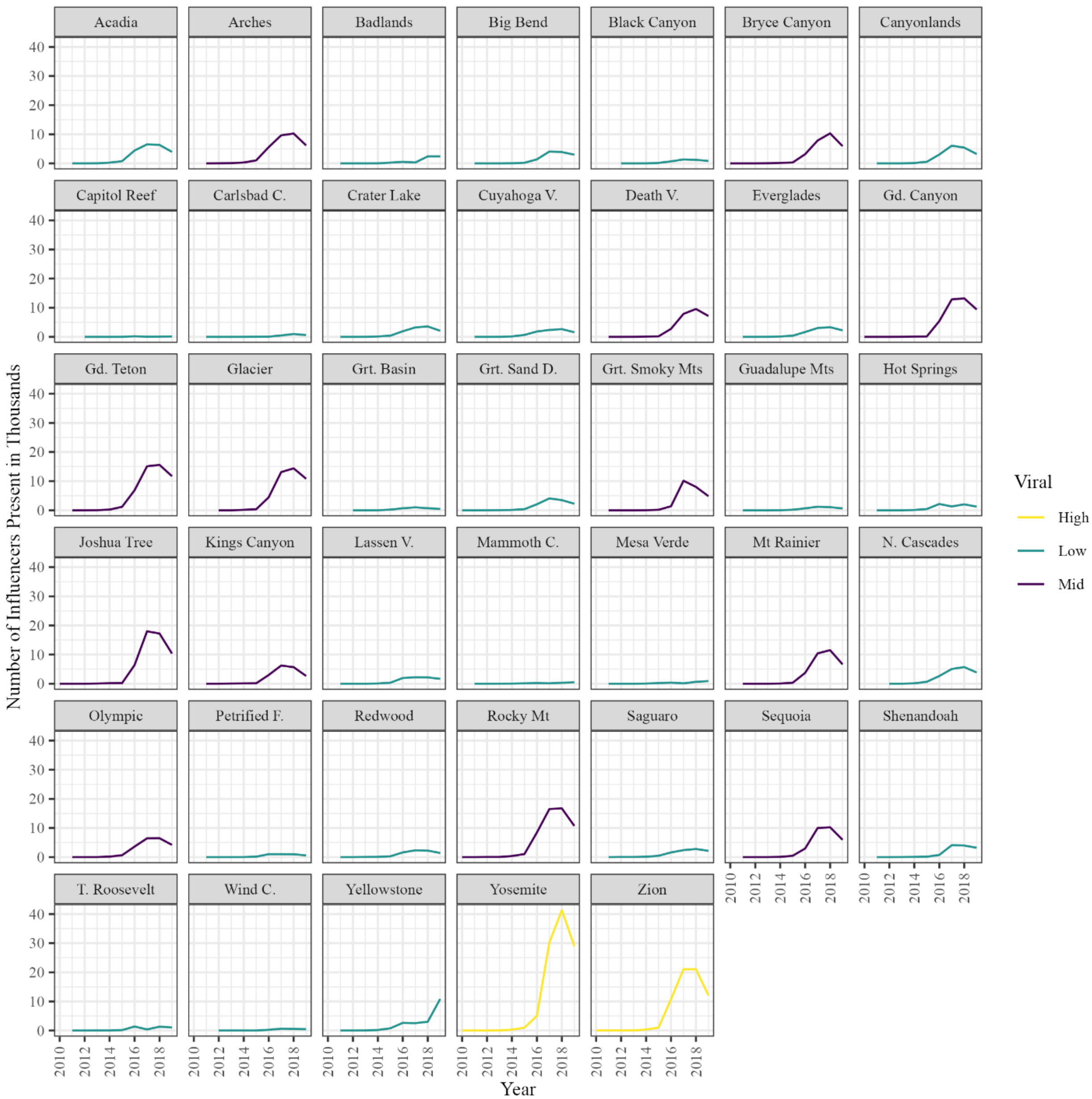
Extended Data Fig. 2: Example of word clouds from Zion and Saguaro National Parks.
Zion Naitonal Park wordcloud of word frequency in captions (top). Saguaro National Park worldcloud of word frequency in captions (bottom). Font size is weighted on frequency. The symbol in Zion is an emoji code.



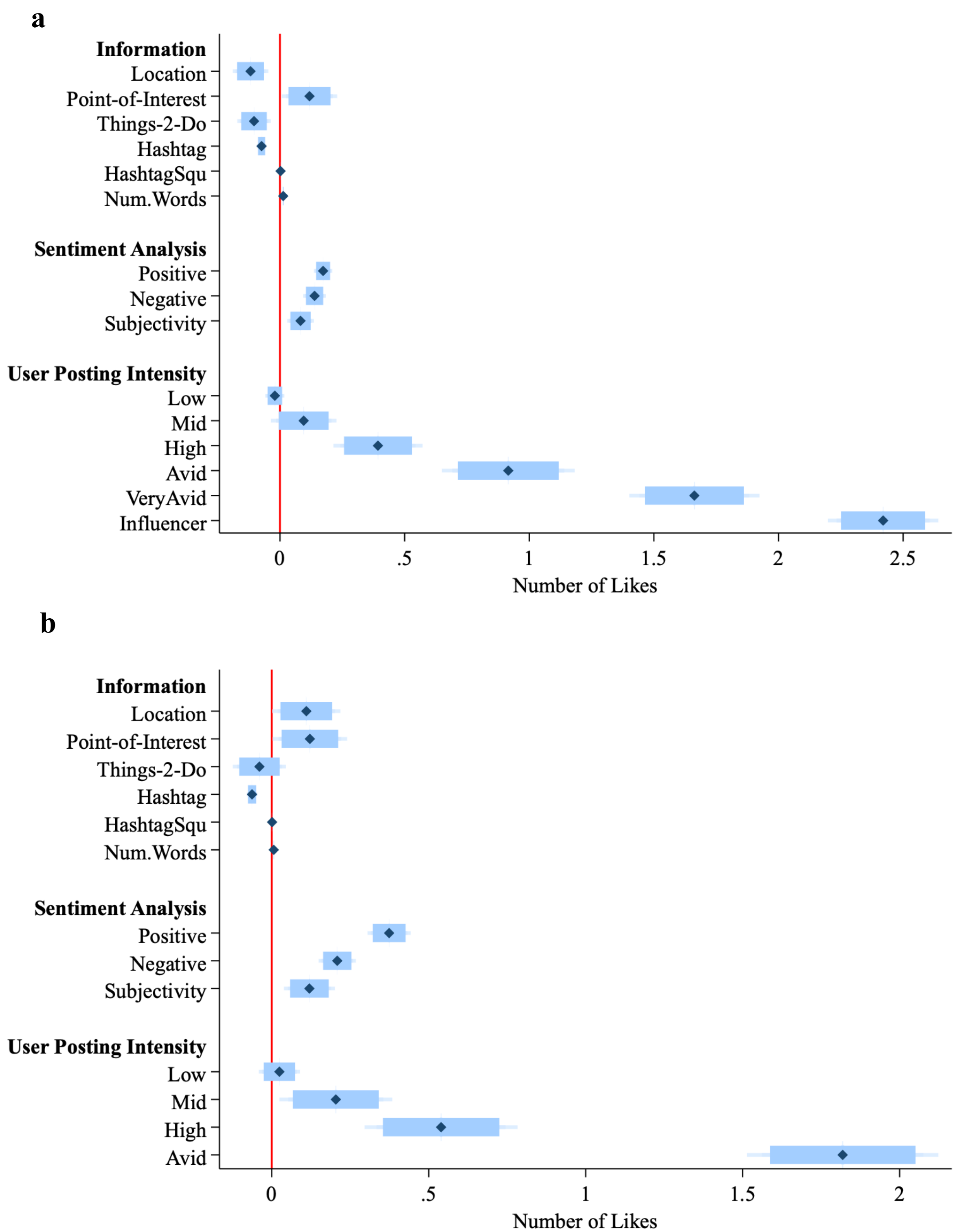
Extended Data Fig. 3: Partial Dependency Plots. The plots are shown for six features of importance (x-axis) from random forest analysis (Fig 3.b) and are used to understand the relationship between variable in the model and the predicted outcome of the number of likes. The number of times a user posts is most important and the upward-sloping path suggest an increase in the number of posts tends to increase the predicted number of likes. Flat regions in the plot indicate that changes in the variable have little influence on predictions within that range. The ticks marks along the x-axis indicate the deciles of a distribution of likes.



Extended Data Fig. 4: Most Viral Photo Based on Likes for Each Park Per Year. Each USNPs most viral photo for each year is plotted based on the number of likes the photo received. Parks grouped in low/moderate/high engagement group are in purple/green/yellow.



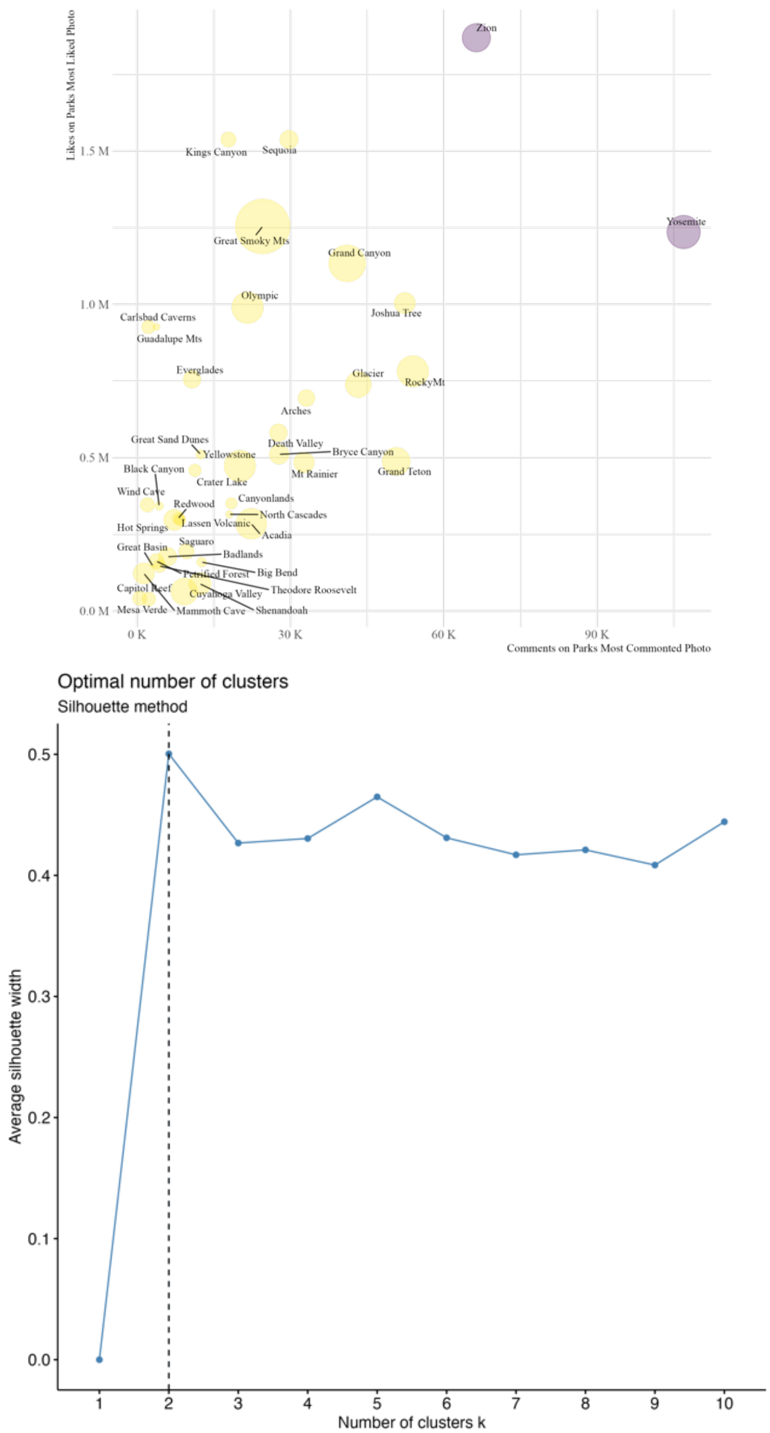
Extended Data Fig. 5: Influencers at Each Park Per Year. Each park's most viral photo for each year is plotted based on the number of avid user (Influencers) observed at each park. Parks grouped in low/moderate/high engagement group are in purple/green/yellow



Extended Data Fig. 6: Robustness Checks on Engagement Model Results. Panel (a) displays results from a negative binomial regression (eq. M.2) with extended user intensity percentiles: 20th, 50th, 75th, 90th, 95th, and 99th. Panel (b) displays results from estimating eq. M.2 as a Poisson regression.

a.

b.



Extended Data Fig. 7: Clustering Results for Social Media Exposure Groupings. Panel (**a**) shows **t**he DBSCAN method that identifies Yosemite and Zion as outliers, driven by their high social media exposure. These parks are excluded from subsequent analysis to better explore the underlying cluster structure. In panel (**b**) for the remaining parks, the optimal number of clusters is first determined using the silhouette method. A k-medians clustering is then applied, resulting in the grouping shown in Figure 4.

Extended Data Table 1: Textual & Sentiment Analysis w/ Natural Language Toolkit VADER

National Park	Caption (Uncleaned)	Location	POI	Things-to-Do	Sentiment Analysis
Acadia	#view of #TheBowl on our #hike up! #amainezing #acadia #acadianationalpark #beautifulday #beehivetrail #ocean #instasky #sundayfunday #maine #newengland #ig_newengland	1	1	1	Neutral
Arches	Day trip to #Moab Delicate Arch is looking fine... #iphoneonly #landscape #nature #mountains #moab #arches	1	1	0	Positive
Badlands	Bad bad lands badlands	1	0	0	Negative
Bryce Canyon	a photo just doesn't do it justice	0	0	0	Negative
Everglades	#everglades #alligators #alligator #dead #vulture #florida	1	1	0	Negative
North Cascades	All smiles on the ridge of Sharkfin Tower a few weeks back, smoky skies and all.	0	1	0	Positive
Yellowstone	another ridiculous sunset in the stone	0	0	1	Negative
Yosemite	Rockclimbing El Capitan	0	1	1	Neutral
Zion	i miss snow & Zion	1	1	1	Negative

Note: Example of textual analysis results from both park-based dictionaries and identifying location, points of interest, things to do and the results of classification based on the sentiment analysis.

Extended Data Table 2: Comparison of Visitation Model Results with Different Standard Error Choices

Model	(1) IV	(2) IV	(3) IV	(4) IV
Standard Errors (SE)	Two-Way Cluster Park-Month & Year-month	Robust	Cluster Park	Two-Way Cluster Park & Year-month
1(post2010)* 1(SMG_H)	0.168*** (0.050)	0.168*** (0.035)	0.168 (0.1189)	0.168 (0.1186)
R^2	0.018	0.018	0.018	0.018
1(post2015)* 1(SMG_H)	0.148** (0.059)	0.148** (0.047)	0.148 (0.159)	0.148 (0.158)
R^2	0.008	0.0008	0.008	0.008

Note: All models have 14,141 observations and include Park, Year, and Month fixed effects and additional control variables. 1(post2010) and 1(post2015) are indicator variables for two different timing options to denote the impact of Instagram. 1(SMG_H) is an indicator variable for a park in the high social media exposure group as defined by Wichman (2024). Col. (1) shows IV results using SE choice from Wichman (2024). Col. (2) shows results are functionally equivalent to using robust SE. Cols. (3) and (4) show that the statistical significance is no longer present when clustering by park or a two-way cluster that does not include month in each cluster. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Extended Data Table 3: Descriptive Statistics for Geotagged Photos to USNPs on Instagram

	Geotagged Photos (N)	*Number of Likes* 90th percentile	*Number of Likes* 95th percentile	*Number of Likes* 99th percentile
Total				
2010-2019	8,310,295	207	347	2,047
Annual				
2010	10	11	12	13
2011	706	22	43	307
2012	9,383	39	67	269
2013	22,723	57	94	346
2014	90,157	95	158	737
2015	215,324	129	223	1,114
2016	1,087,549	163	284	1,460
2017	2,267,731	211	371	1,885
2018	2,462,000	227	425	2,436
2019	2,153,712	216	395	2,304

Extended Data Table 4: Descriptive Statistics for Influential Posts and Avid Users

	Full Data	90th percentile	95th percentile	99th percentile
Number of Unique Users	2,790,954	10,868	2,953	170
Number of Posts	8,310,295	41	79	272