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

Ashley C. Lowe Mackenzie
Department of Natural Resources and Environmental Management
University of Hawai'i at Mānoa
Honolulu, HI 96822, USA
alowemac@hawaii.edu

Steven J. Dundas
Department of Applied Economics & the Coastal Oregon Marine Experiment Station
College of Agricultural Sciences
Oregon State University
Corvallis, OR 97331, USA
steven.dundas@oregonstate.edu

Abstract:

Recent increases in recreational visits, staffing shortages, and maintenance backlogs have left public land agencies in the United States facing multiple stressors that affect their ability to manage land sustainably. One hypothesis posited by traditional media outlets is that social media is a primary reason for recent visitation increases and resource degradation on public lands. We test the validity of such claims and investigate how online content and behavior on Instagram may connect to visitation trends at US National Parks. Using millions of georeferenced Instagram posts, we find parks experienced visitation increases attributable to Instagram but the likely reasons vary across parks. A few locations saw increases associated with the cumulative impact of viral content while small increases at all other parks were associated with the volume of geotagged content. Exploration into potential mechanisms suggests that viral posts by avid users may influence visits at a broader set of parks.

Management of the nearly 640 million acres of public land in the United States (US) has multiple objectives, including balancing the conservation of species and ecosystems with recreational use. Trends in recreational visits to U.S. National Parks (USNPs) have experienced multiple periods of growth over the last century (**Fig. 1**). There was a steady increase in visitation following the end of World War II that lasted for decades[1,2]. This trend is attributed to population growth and rising levels of income, along with improvements in travel and accessibility[1,2]. By the early 2000s, visitation trends were stagnant, leading some researchers to investigate reasons for the lack of growth[3,4]. This period was short-lived and USNPs began reporting record-breaking annual visitation beginning in 2013. This recent visitor overcrowding adds stress to staff, facilities, ecosystems and aging infrastructure[5,6]. This latter point is significant as the US National Park Service (NPS) currently faces a $22.3 billion dollar backlog of deferred maintenance[7].

A common culprit often portrayed in traditional media outlets and land management circles for the recent rise in visitors (and the resulting resource degradation) to USNPs is the social media platform Instagram[8–11]. Instagram began as a photo-sharing app and has shouldered blame for the growing crowds given the timing of the app's release, the nature of the content (e.g., georeferenced images), the way content is delivered to its 1 billion active monthly users and reports by park managers reporting swarms of visitors to particular locations observed taking self-portraits, or selfies.[10–12] Data from Instagram and other social media platforms have also been shown to provide practical benefits to land managers. Research has found social media data is a reliable proxy to understand the number of visitors to public lands[13–17]suggesting avenues to reduce costs of data gathering across managing agencies and extend visitor use estimates to wilderness areas where prior data collection was limited[13,15,18–21]. Others have found that National Parks location which have created online accounts on sites such Twitter and Instagram have a causal relationship with an increase in visitation in the 2010s[22]. Linking within-app behavior to visitation has been less studied,[22] talthough 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[12].

The USNPs are 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. Investigation of the influence of Instagram on visits to USNPs and the potential mechanisms driving in-app engagement may reveal important insights to improving sustainable management of these iconic places. Considering how the potential influence of Instagram is driven by activity from the user base within the app (**Fig. 2**), we have two hypotheses about the link between visitation increases and Instagram. Higher visitation could be driven by 1) a reduction in search and information costs[23] provided by the app in helping people find recreational opportunities or it could be 2) 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[24].

This study analyzed over 8.3 million public posts from Instagram georeferenced to USNPs with a set of visitation models to examine if and how the app has influenced visitor trends to these well-known locations. We leverage the timing of Instagram's rise to over a billion monthly active users, park-specific contemporaneous upload activity and user engagement with content, and the information provided within this content posted to Instagram to determine the level and location of visitation changes and identify potential mechanisms for those impacts. Our results show all USNPs likely experienced an increase in visitation after Instagram gained influence, but the reason for the increases vary across parks. We estimate that USNP locations that had "viral moments," where large numbers of users interacted with georeferenced content, saw an increase attributable to the cumulative impact of all current and past park-specific viral content (a potential bandwagon effect). Results from USNP units without similar viral content show small visitor increases associated with the volume of contemporaneous upload activity (a potential indication of reduced information costs). To further explore potential mechanisms for these changes, we then use text data from the captions of each post and posting frequency from each user to determine what types of information and in-app actions are associated with increased user engagement. Our results suggest that specific locational information (Points-of-Interest), expressions of both positive and negative sentiment about the experience and content from high frequency posters to Instagram all increase user engagement with content. We use these additional findings to show visitation to parks with the highest viral moments were likely impacted by viral content posted by any user type while a second group of parks with moderate viral moments likely experienced increased visits attributable only to high frequency, or avid, users who post georeferenced content to those parks repeatedly.

Results

New name

We selected 40 USNP located in the conterminous United States that have ease of vehicle accessibility for the average visitor (**Methods**). We compare how our park selection impacts results of the previous results of social media influence on visitation[25]. First, our inclusion of controls such as population, demographic factors, trends in gasoline prices and fees impact the overall magnitude in the relationship of social medias impact in visitation. We also find by excluding parks which require vast travel to reach for the average visitor, such as Alaska and America Samoa, instruments to controls for trends in parks popularity located in the contingent United States prior to social media suggest the effect is undetectable. We show that the app's launch in October 2010 does not systematically change visitation at USNPs (**Fig. 3.A**). Other important factors affecting visitation in this baseline model included the percentage of the US population over 65 years of age, extreme hot or cold temperatures, state population and the unemployment rate (**Fig. 3.B)** and the facts that it corresponded with Instagram being purchased by Facebook (currently known as Meta), an expansion of the app to the Android operating system and the app itself reaching 50 million daily active users. We show that all National Parks within our study experienced a significant increase in average visitation (~ 10 percent) after April 2012 relative to before this date (**Fig. 3.A**).

This estimate suggests there was a change in visitation trends to USNPs in the spring of 2012 and to more clearly identify how much of this effect is actually attributable to Instagram, we add park-specific data to the model that captured monthly counts of georeferenced photo-user-days (PUDs)[12] representing the volume of daily user activity on the app. We found the number of PUDs is associated with a small, marginally significant increase in overall visitation (**Fig. 3.C**) to all parks. This specification only controls for the contemporaneous act of uploading content to a park in a given month and assumes all USNPs are experiencing similar activity levels from visitors within the app. However, 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, suggesting further exploration was warranted.

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 an average georeferenced PUD. 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.

To evaluate viral content in our context, we plot each USNP unit based on the number of likes and comments of the most viral georeferenced content in our dataset (**Fig. 4**). There is significant variation among parks based on their most viral moment and two parks are clear comparative outliers on Instagram - Yosemite and Zion NPs. Sixteen additional parks create a second cluster in **Fig. 4** that suggest moderate viral moments while the remaining twenty-two parks in our sample had visible engagement but lower viral moments on Instagram compared to other USNPs. We plot the average visitation trends of the parks in the high, moderate, and low viral group described above from 1993 to 2019 (**Fig. 5**) and the outcome 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. The high viral parks show the largest increase while visitation at the parks in the low viral group appears relatively constant across the 26-year period. Delineating our PUD regression finding an average 10 percent rise in visitation by viral grouping, we find that contemporaneous PUD activity does not affect visitation at park units in the high and moderate viral groups and the impact is driven by PUDs at low viral group parks (**Fig. 3.C**). This is suggestive of an information effect as visitation increases for a majority of parks are correlated with contemporaneous georeferenced content.

We next add a variable to the model to test whether content with high viral moments (e.g., engagement) impacted visitation to USNPs. This content is subset based on individual posts that receive a number of likes in the 90th and 95th percentiles each year for all content uploaded using a USNP geotag **(Extended Data Table 1)**[12]. We then quantify a cumulative viral effect for each park every month as the aggregate of both past and current viral content. When added to our visitation model, we found evidence that this cumulative viral effect increased visitation at the two parks in the high viral group by 7.7% per month (**Fig. 3.D**). This positive impact of visitation is also present for parks in the moderate viral group for viral content in the 95th percentile. We find no significant effect of cumulative viral content on visitation at USNP units in the low viral group.

These results suggest the potential for a bandwagon effect at parks with high to moderate viral exposure on Instagram.

Unpacking Engagement with USNP Instagram Content

To further test our hypotheses about specific mechanisms driving our visitation results, we processed all text content of each of the 8.3 million geotagged Instagram post 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 (the state the park is located or the park name), points-of-interest listed by USNPs managers as important locations within the park and/or the activities listed as advertised *Things-To-Do* on each individual park's website.[26] We considered the extent of information provided by controlling for both the number of words within the caption and the number of hashtags used and conducted a sentiment analysis to control for positive, neutral or negative language expressed in each post (**Methods**). The subjectivity of the each caption was analysized to capture whether the content contained more opinions, emotions, or personal interpretations or more objective and factual information. We identify 2.8 million unique users posting content with USNP geotags and classify these users by summing the number of times an individual uploads content to any National Park geotags over our sample frame after the introduction of Instagram. We group users of the app by quantile following a similar strategy used to identify viral content for the visitation models (**Methods**). We define those posting content in the 90% percentile of users 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, 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 all the above information in a negative binomial count data model to determine how these elements affect the number of likes (engagement) an Instagram post is likely to receive from the general user base. We found user posting intensity (avid users) had the largest effect on the amount of engagement recieved from online users (**Fig 6.A**). Avid users are expected to have an engagement rate 4.5 times greater than all other users. This result provides supportive evidence the mechanism driving engagement is more of a behavioral effect (e.g., bandwagon) than a pure information shock. That said, informational content contributes to an increase in engagement as park-specific points-of-interest information increased engagement by 1.1 times more than content from users not discussing specific locational information. The results show that the number of hashtags 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 2**). Lastly, we show that opinion-based informaiton also had a positive effect on engagement relative to factual information.

To validate these findings, we conducted a robustness check using a random forest model to estimate the relative importance of each covariate in our model (**Fig 6.B**). Results from the feature importance scoring found that user posting intensity is the most important feature in understanding the level of engagement the content recieves. This can be seen clearly in the path dependency of the features within the the random forest model **(Extended Data Figure 3).**

Given this finding of the importance of user posting intensity, we then re-estimated two subsets of our visitation model with PUDs and cumulative viral impacts across park groups – one using only posts from avid users and the other with all content from all other users. Importantly, avid users are responsible for 27% of all viral content. This analysis provides complimentary findings from our primary visitation model (**Figure 3.D**) with one main difference. These new estimates suggest viral content from avid users significantly increases visitation at high and medium viral parks at higher rates than the original model that did not account for user posting intensity (**Fig 6.C**). When we examine the remaining 73% of the viral content from non-avid users, the positive impact on visitation only holds for the two parks in the high viral group (Yosemite, Zion). These findings suggest two USNP units are likely to have seen visitation increases associated with viral content from all user types and parks in the moderate viral group were likely impacted by viral content from specific user types (avid posters). The cumulative effect per month of all past and current engaging content from avid users increases visitation by 4.4 to 7.2% per month across parks in the high and moderate groups. To summarize, moderately viral content from avid users is likely to influence visitation but highly viral content has impacts regardless of the source of the post. USNPs units in the group with relatively low viral content do not appear to have any visitation impacts associated with either the cumulative effect of such content or user posting intensity.

Discussion

Online activity has offline impacts[27] and visitation to USNPs is no exception. New information from social media has the potential to make the outdoors less exclusive and increase the diversity of recreators[28] yet the conversation has often focused largely on its negative role in driving surges of visitation, leading to overcrowding and resource degradation. Our analyses that combine visitation models with data from millions of posts to Instagram suggest that part of the increased visitation since 2012 at USNPs may be attributable to online behavior within the app. 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.

Importantly, our analysis shows that Instagram content alone does not fully explain the overall rise in visitation observed. We found that demographic trends also played a large role in the recent increases in visitation. Specifically, the national population over the age of 65 increased rapidly over this same time period and accounted for a large portion of the overall increase visitation. 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 in influencing park visitation trends. In other words, our evidence suggests some impact from Instagram that varies by park unit, but the app itself is not solely responsible for the record-breaking increased visitation to USNPs in the last decade as some traditional media outlets have claimed[7-10].

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 avid accounts producing high engagement content that is likely impacting visitor patterns in specific areas. Our results suggest that park managers could monitor social media posts tagged to their location and be more prepared for visitation increases if content associated with their location has a viral moment. 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[27,29]. Integrating pertinent information about location-specific requirements or initiatives such as "Take only photographs, leave only footprints" directly within social media apps[30] 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 or engagement indicators, 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.[31] The database contains information on visitor trips to over 400 park locations managed by the NPS. The type of a 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 areas are removed since they 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 therefore a monthly panel covering a nearly 27-year period for 40 NPs.

Social Media Data

Instagram data was collected by web scrapping publicly available content in fall of 2019. We conducted this effort following the processes described in Lowe Mackenzie et al. (2024)[11]. This collection effort of all geotags associated under each NP in our sample resulted in over 8.3 million geotagged posts to these locations. 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 protect 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)[12] 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* and *comments* a photo receives. We measured the cumulative effect of Instagram posts with high engagement using the following equation:

$$Infl_{pT} = \sum_{t=1}^{T} Content_{pt}^{l_y} \ , \tag{1}$$

where $Infl$ is the cumulative (T) influential posts for park p, 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 threshold, l, of the number of likes a post receives. The threshold is determined if the post is in the top 90th and 95th based on the total number of likes received on the entire set of photos geotagged per year (**Extended Data Table 1**).

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 viral content based on the most commented and most liked photo. **Fig. 4** visually suggested potential grouping strategies based on viral moments. The y-axis position marked the number of likes for the highest-liked content for each park and the x-axis marked the highest number of comments on any content for each park. The high viral grouping includes Yosemite and Zion NPs. The moderate viral group contained sixteen (16) parks: Kings Canyon, Sequoia, Grand Canyon, Joshua Tree, Great Smoky Mountains, Everglades, Death Valley, Bryce Canyon, Carlsbad, Guadalupe Mountains, Glacier, Grand Teton, Mount Rainier, Olympic, Arches and Rocky Mountain. The remaining twenty-two (22) parks had minimal viral content and were considered the low viral group. **Extended Data Figs. 3-4** shows the likes and comments over time of content at each park.

$$InfPost_{pT}Content_{pt}^{l} \ ly. l,$$

Other Covariates

The structure of our data (monthly panel) necessitated the inclusion of additional covariates at the same time step. Macroeconomic variables that may influence visitation trends were retrieved from the Federal Reserve Bank of St. Louis[32]. We collected a seasonally adjusted national monthly income level measuring real disposable personal income (RDPI)[33] and the state-level unemployment rates, which captured a measure of consumer confidence[4,32]. State population totals are available as an annual estimate[32,34] 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 was from the home state of the park unit. 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[4,35]. 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)[36]. Lastly, weather conditions may also affect visits to recreation sites[36]. Daily observations of maximum temperature and precipitation for all parks were collected from PRISM Climate Group at Oregon State University[37]. Exploiting this daily variation, weather enters our model as a non-linear response function using a binning approach[38,39]. 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 the 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 reacting to the temperature maximum rather than the temperature mean. Another spatial concern arises given the national scope of this project. Locational variance in temperature is heterogenous across regions. Deserts areas can have higher highs as well as very cold nights – extremes that could get lost in an average measure. In such a case, a desert and a forested area with milder weather may appear to have a similar mean temperature distribution but a maximum temperature approach could account for and use the higher variance in deserts.

Visitation Models

The baseline model specification examined the impact of Instagram's launch on visitation at park, *p,* in month t ($Visit_{pt}$) using a binary indicator to denote the pre- and post-Instagram periods, IG_{Launch_t} , and a second version of the model refined the timing to when Instagram had gained influence in the social media space (April 2012, IG_{Use_t}). The dependent variable is logged to reduce the impact of outliers and ease interpretation of model coefficients as percentage changes or elasticities. The baseline model was specified as follows:

$$Ln\big(Visit_{pt}\big) = \beta_0 + \beta_1 IG_{Launch_t} \boldsymbol{or}\ IG_{Use_t} + \beta_2 \mathbf{T}_{pt} + \beta_3 \mathbf{PRE}_{pt} + \beta_4 \mathbf{X}_t$$
$$\beta_5 Ln(Fee)_{pt} + \beta_6 Desig_{pt} + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \ . \qquad (2)$$

In eq. 2, $\mathbf{T}_{pt}$ represents daily counts of days per month in observed 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) at time *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, $\rho_p, \tau_{y(t)}, \gamma_t$ control for park, year and month fixed effects, respectively. The error term is ε_{pt} and all models are estimated with robust standard errors.

To examine the role of general upload activity, we added PUDs[13] by park/month to the baseline model:

$$Ln(Visit_{pt}) = \beta_0 + \beta_1 IG_{Launch_t} + \beta_2 PUD_{pt} + \beta_3 \mathbf{T}_{pt} + \beta_4 \mathbf{PRE}_{pt} + \beta_5 \mathbf{X}_t$$
$$\beta_6 Ln(Fee)_{pt} + \beta_7 Desig_{pt} + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \ . \quad (3)$$

We used IG_{Launch_t} instead of IG_{Use_t} because PUDs are possibly non-zero once the app is launched. The final specification added the cumulative effect of viral content at each park (eq. 1) to the model. This is to determine if and how content with high levels of user engagement impacted visitation over time and through a different channel than contemporaneous upload activity.

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

This specification also adds subscript G to the *PUD* and *Infl* terms to reflect the high, moderate, and low 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.

Textual and Sentiment Analysis of Captions

We found certain parks receiving a high level of engagement on Instagram experienced cumulative impacts from highly engaging content, indicative of a potential bandwagon effect. For parks with relatively lower viral moments, we found that contemporaneous upload activity was associated with visitation increases, suggestive of a reduction in search costs. To explore potential mechanisms further, 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 observation includes 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 assess 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. 4).** 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 parks 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. 4).** 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 valuable 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[42]. 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 colloquialism, intense punctuation (!!), all caps emphasis (this is VERY awesome), emoticons (☺), as well as acronyms commonly used in online slang such as LOL. The process estimates the intensity of sentiment to determine whether the statement has an overall positive, negative, or neutral tone. **Extended Data Table 2** provides a sample of the sentiment analysis results of all park captions. 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.

Estimating Factors that Influence Engagement

Our park-specific data dictionaries and sentiment and subjectivity analyses provide further 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[43].

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

$$Likes_{pi} = \beta_1 Location_p + \beta_2 POI_p + \beta_3 Things2Do_p + \beta_4 Sentiment_i + \beta_5 Subjective_i + \beta_6 Hashtags_i + \beta_7 Hashtags_i^2 + \beta_8 Num.Words_i \; \beta_9 UserInt_{pi}^Q + \rho_p + \tau_{y(t)} + \varepsilon_{pt}\,, \quad \text{(M.1)}$$

where $Location_p$,POI_p, $Things2DO_p$ are binary indicators that are equal to one if the post contains text included in the three park-specific dictionaries, respectively. $Sentiment_i$ is a categorical variable using a determination from VADER on whether the caption expressed sentiment that was considered positive, negative or neutral. $Subjective_i$ is a value between 0 and 1 determined by Textblob capturing the extent of fact or opinion within the text. $Hashtags_i$ 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. $UserInt_{pi}^{Q}$ is 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 observe influencers or monetized accounts), we captured this variable as posting frequency at NP 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 25^{th} to 50^{th} percentiles, while moderately active users post frequency falls between the 50^{th} to 75^{th} percentiles. High activity users post frequency falls between the 75^{th} to 90^{th} percentiles. We classified "avid users" if their posting behavior fell in the top 90th quantile. ρ_p, $\tau_{y(t)}$, and γ_t represent park, year (y) and month fixed effects.

Unsurprisingly, due to the very large size of the 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[44,45]**.** 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[46]. The results of this feature importance supported 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)[45] are provided for interpreting the predictions of the random forest model and demonstrate the robustness of our findings (**Extended Data Fig. 5)**.

Our final model incoporated these findings from our engagement analysis into **eq. 4**. 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. 1** with the 90th percentile threshold. We compared results from our original subset of the 90th percentile of engagement (**Fig. 3.D**) to results when differentiating engaging content coming from avid and non-avid users. This comparative excercise 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.

References

1. Clawson, M. Outdoor recreation: Twenty-five years of history, twenty-five years of projection. *Leis. Sci.* **7**, 73–99 (1985).
2. Weiss, T. Tourism in America before World War II. *J. Econ. Hist.* **64**, (2004).
3. Stevens, T. H., More, T. A. & Markowski-Lindsay, M. Declining National Park Visitation: An Economic Analysis. *J. Leis. Res.* **46**, 153–164 (2014).
4. 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).
5. Timmons, A. L. Too Much of a Good Thing: Overcrowding at America's National Parks. *Notre Dame Law Rev.* **94**,.
6. 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).
7. NPS. Deferred Maintenance and Repairs - Infrastructure (U.S. National Park Service). https://www.nps.gov/subjects/infrastructure/maintenance-backlog.htm.
8. Solomon, C. Is Instagram Ruining the Great Outdoors? *Outside Online* https://www.outsideonline.com/culture/opinion/instagram-ruining-great-outdoors/ (2017).
9. Simmonds, C., McGivney, A., Wilkinson, T. & Canon, G. Crisis in our national parks: how tourists are loving nature to death. *The Guardian* (2018).
10. Hegyi, N. Instagramming Crowds Pack National Parks : NPR. https://www.npr.org/2019/05/28/726658317/instagramming-crowds-pack-national-parks (2019).

11. 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.
12. 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.
13. 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).
14. 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).
15. 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).
16. 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).
17. Wood, S. A. *et al.* Next-generation visitation models using social media to estimate recreation on public lands. *Sci. Rep.* **10**, 15419 (2020).
18. 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).
19. Walden-Schreiner, C., Leung, Y.-F. & Tateosian, L. Digital footprints: Incorporating crowdsourced geographic information for protected area management. *Appl. Geogr.* **90**, 44–54 (2018).

20. 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).
21. Ghermandi, A. & Sinclair, M. Passive crowdsourcing of social media in environmental research: A systematic map. *Glob. Environ. Change* **55**, 36–47 (2019).
22. 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.
23. Stigler, G. J. The Economics of Information. *J. Polit. Econ.* **69**, 213–225 (1961).
24. Leibenstein, H. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *Q. J. Econ.* **64**, 183 (1950).
25. Wichman, C. J. Social media influences National Park visitation. *Proc. Natl. Acad. Sci.* **121**, e2310417121 (2024).
26. NPS.gov Homepage (U.S. National Park Service). https://www.nps.gov/index.htm.
27. Allcott, H. & Gentzkow, M. Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **31**, 211–236 (2017).
28. Flores, D. & Kuhn, K. Latino Outdoors: Using Storytelling and Social Media to Increase Diversity on Public Lands. *J. Park Recreat. Adm.* **36**, (2018).
29. MacCarthy, M. Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.3615726 (2020).

30. 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.
31. STATS - Welcome to Visitor Use Statistics. https://irma.nps.gov/Stats/.
32. Federal Reserve Economic Data | FRED | St. Louis Fed. https://fred.stlouisfed.org/.
33. 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).
34. Bureau, U. C. Data. *Census.gov* https://www.census.gov/data.
35. 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.
36. Homepage - U.S. Energy Information Administration (EIA). https://www.eia.gov/.
37. PRISM Climate Group at Oregon State University. https://prism.oregonstate.edu/.
38. 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).
39. 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).
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. 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).

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

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

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

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

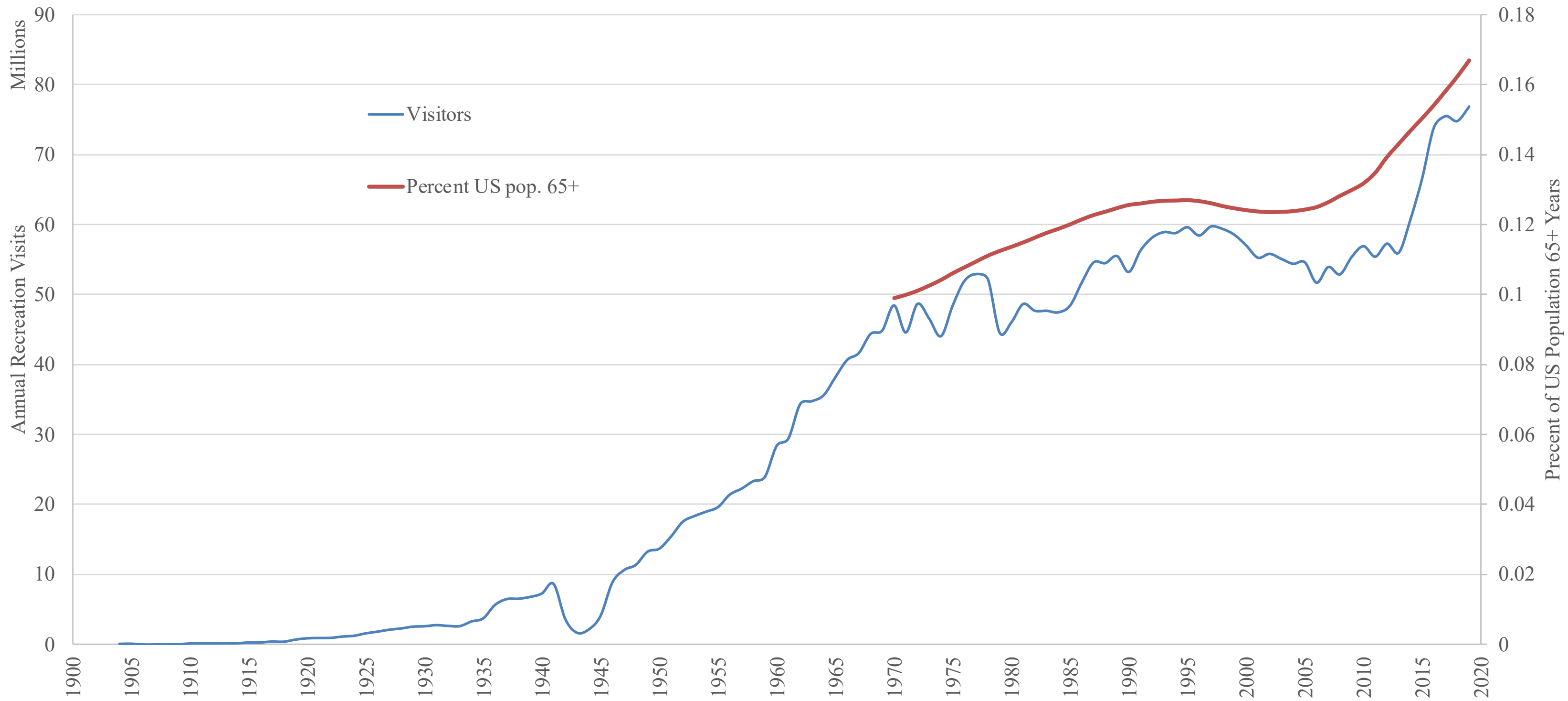


Fig. 1 | Annual recreational visits at U.S. National Parks 1904 to 2020. Left y-axis scale is total annual recreation visitors in millions and the right y-axis is the percentage of the US population 65 or older.

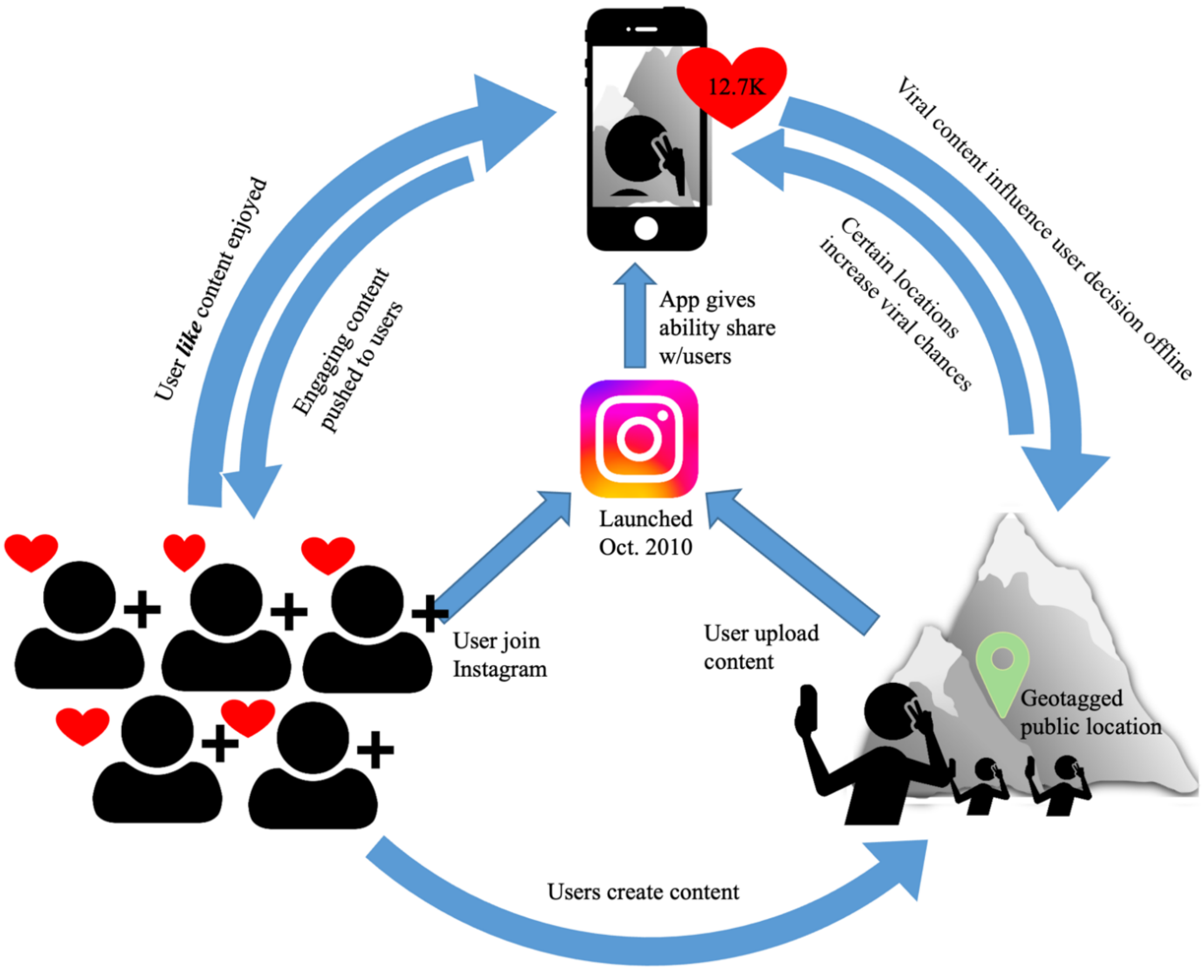


Fig. 2 | Conceptual diagram of Instagram's influence of visitation to public lands. When social media platforms begin, it takes time to develop a user base that uploads content and engages with other users' content. Instagram launched in October 2010 and reached 50 million users by April 2012. As content 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.

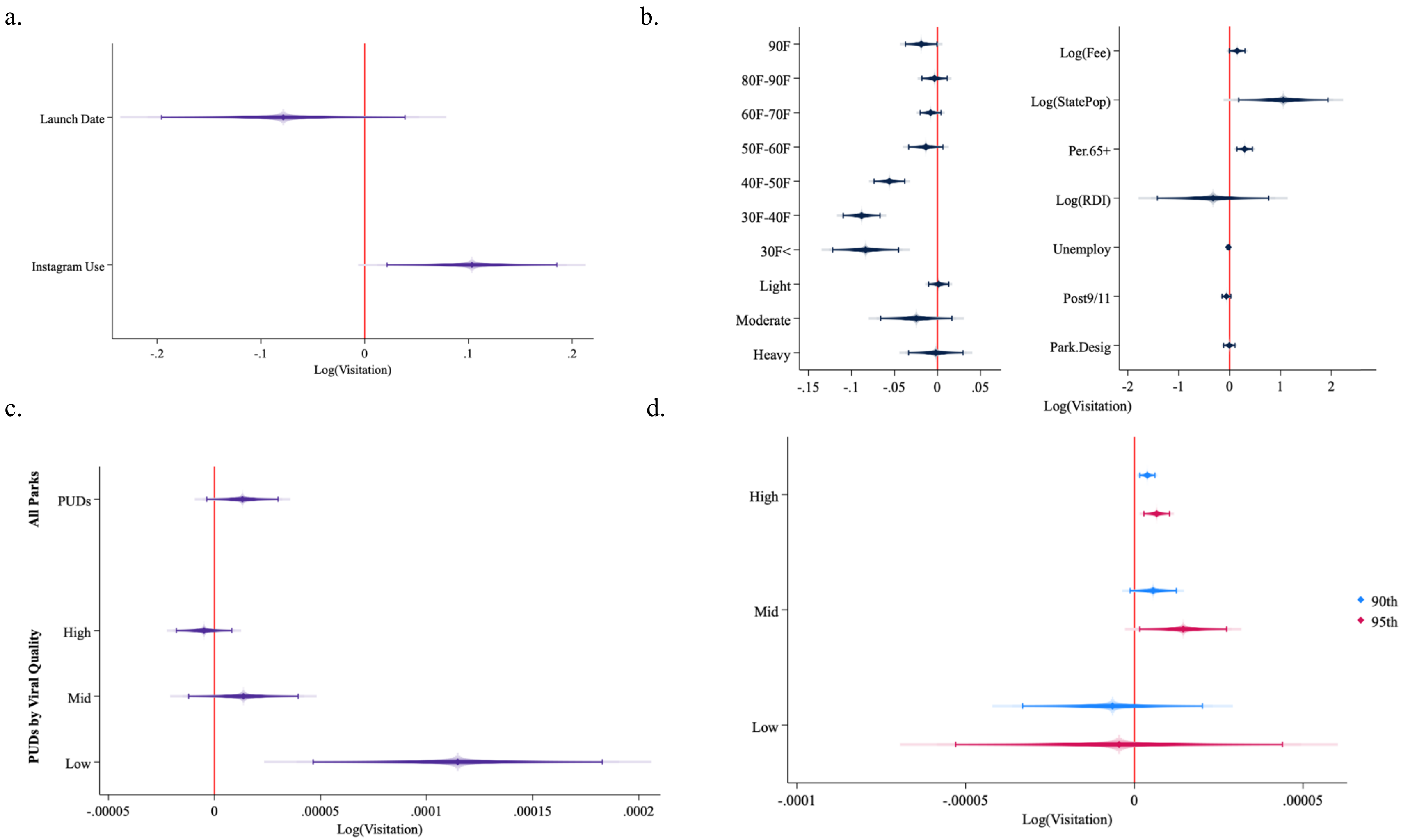


Fig. 3 | Factors influencing visitation trends. Panel (**a**) examines the effect of the timing of Instagram on USNP visitation. Panel (**b)** plots all covariates estimated as controls for our baseline visitation model (**eq. 2**). Panel (**c)** plots the contemporaneous effect of PUD per month for all parks (**eq. 3**) and through groups determined by viral moments plotted in **Fig. 4**. Panel (**d)** plots impact of the cumulative impact of engaging content by park group (**eq. 4**).

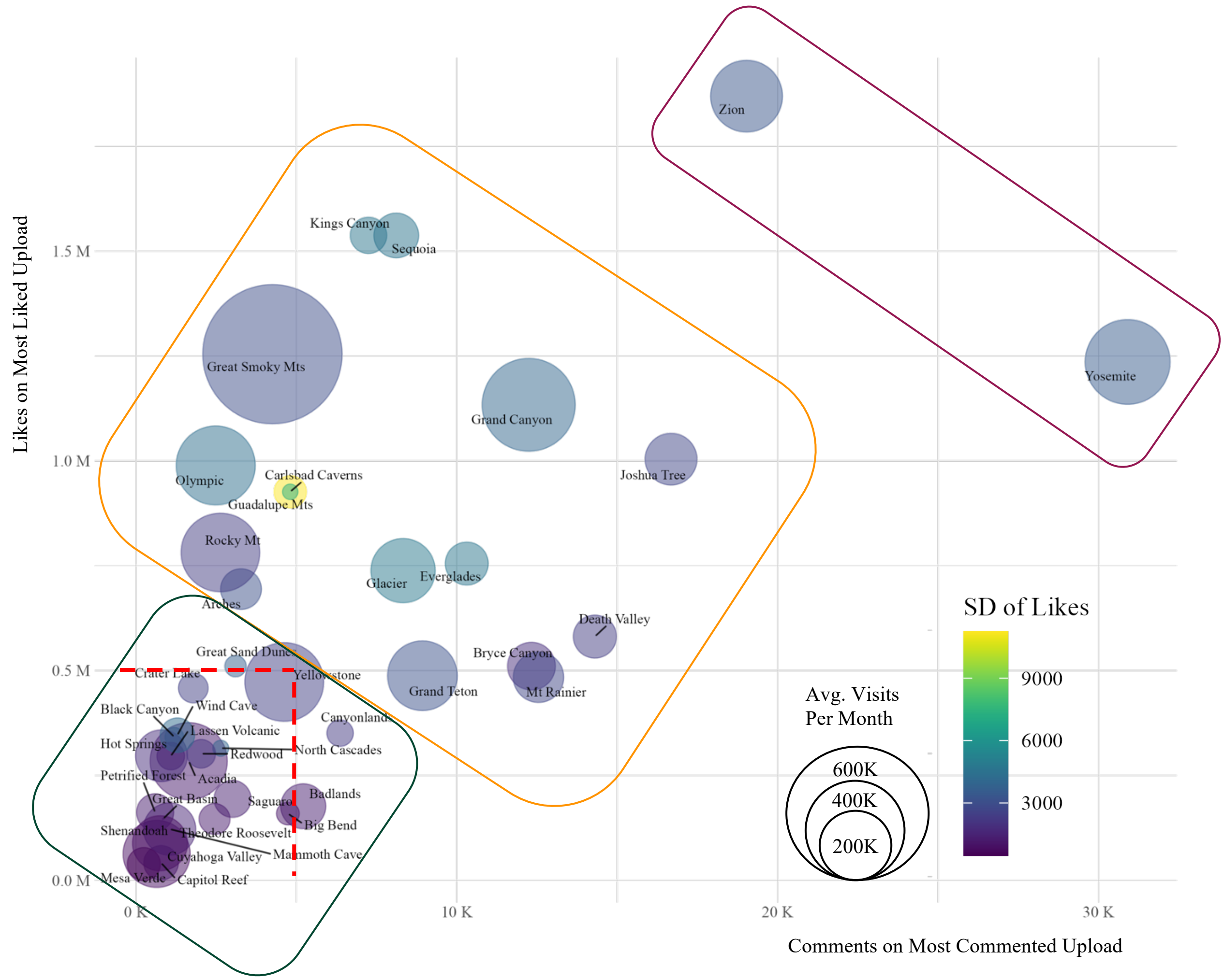


Fig. 4 | Modeling viral moments and visitation at US National Parks 201 - 2019. Y-axis is the number of likes that the most liked photo from each park received. X-axis is the number of comments the most commented photo from each park received. The size of the circle indicates average monthly visitation to park. Color represent the standard deviation (SD) of the number of likes a park recieves on all photos uploaded. Three grouping strategies are suggested in this plot. A High Engagement Group (red rectangle) had the most viral content on the app (Zion & Yosemite NPs). A Middle Engagement Group (orange rectangle) contains parks with less viral content than Zion or Yosemite, but still higher than many parks (Kings Canyon, Sequoia, Grand Canyon, Joshua Tree, Great Smoky Mountians, Everglades & Death Valley, Olympic, Carlsbad Caverns, Guadalupe Mountains, Glacier, Grand Teton, Bryce Canyon, Mount Rainer, Rocky Mountain and Arches NPs). A Low Engagement group (green rectangle) contains all other parks in our study area. Dotted red line represents the same thresholds represented in **Extended Data Figs. 3-4** (500,000 likes and 5,000 comments).

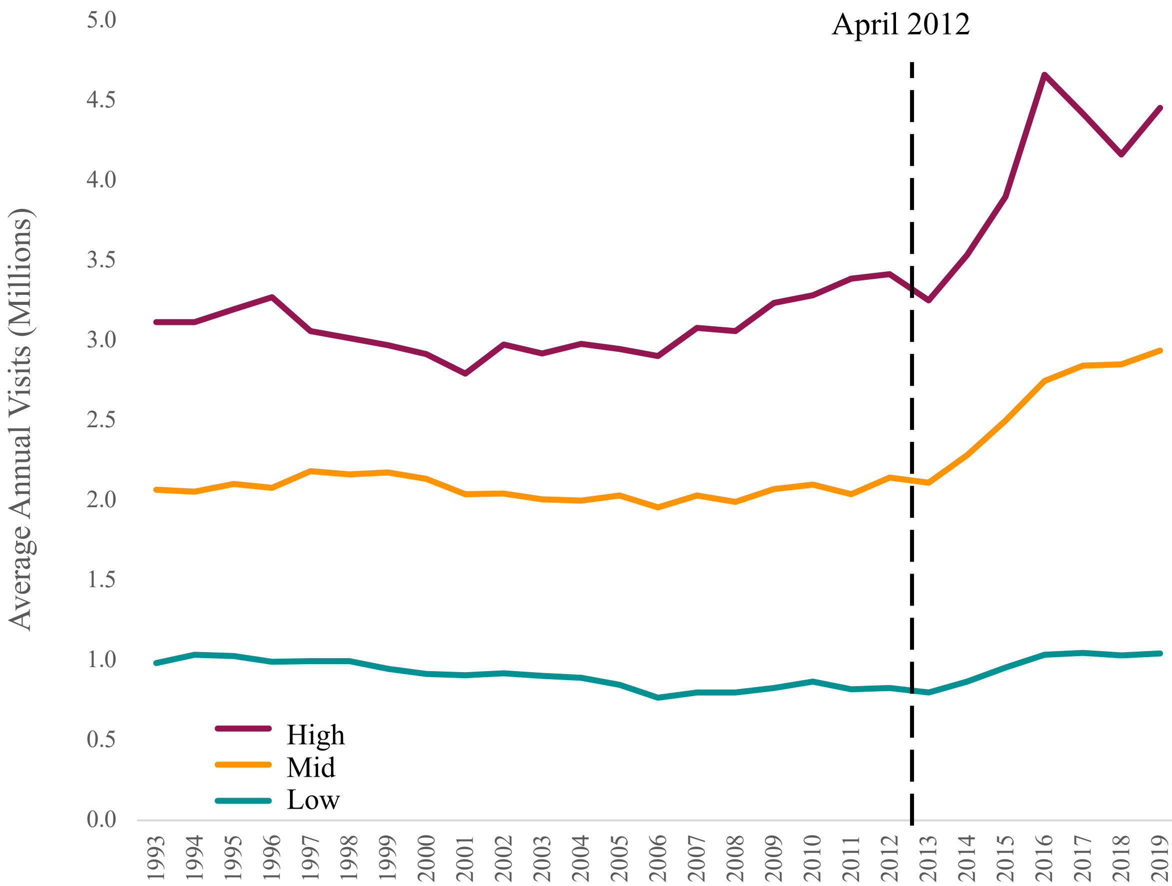


Fig. 5 | Visitation growth by park engagement grouping. Y-axis is the average annual visits per engagement group. X-axis is the time period covered in the analysis. The high engagement group sees more average visitation than the other groups and saw a sharp increase post Instagram. The middle engagement group also shows modest growth after Instagram gained a wide user base. The trend in the low engagement group was realtively steady over 26 year timeline, including after Instagram.

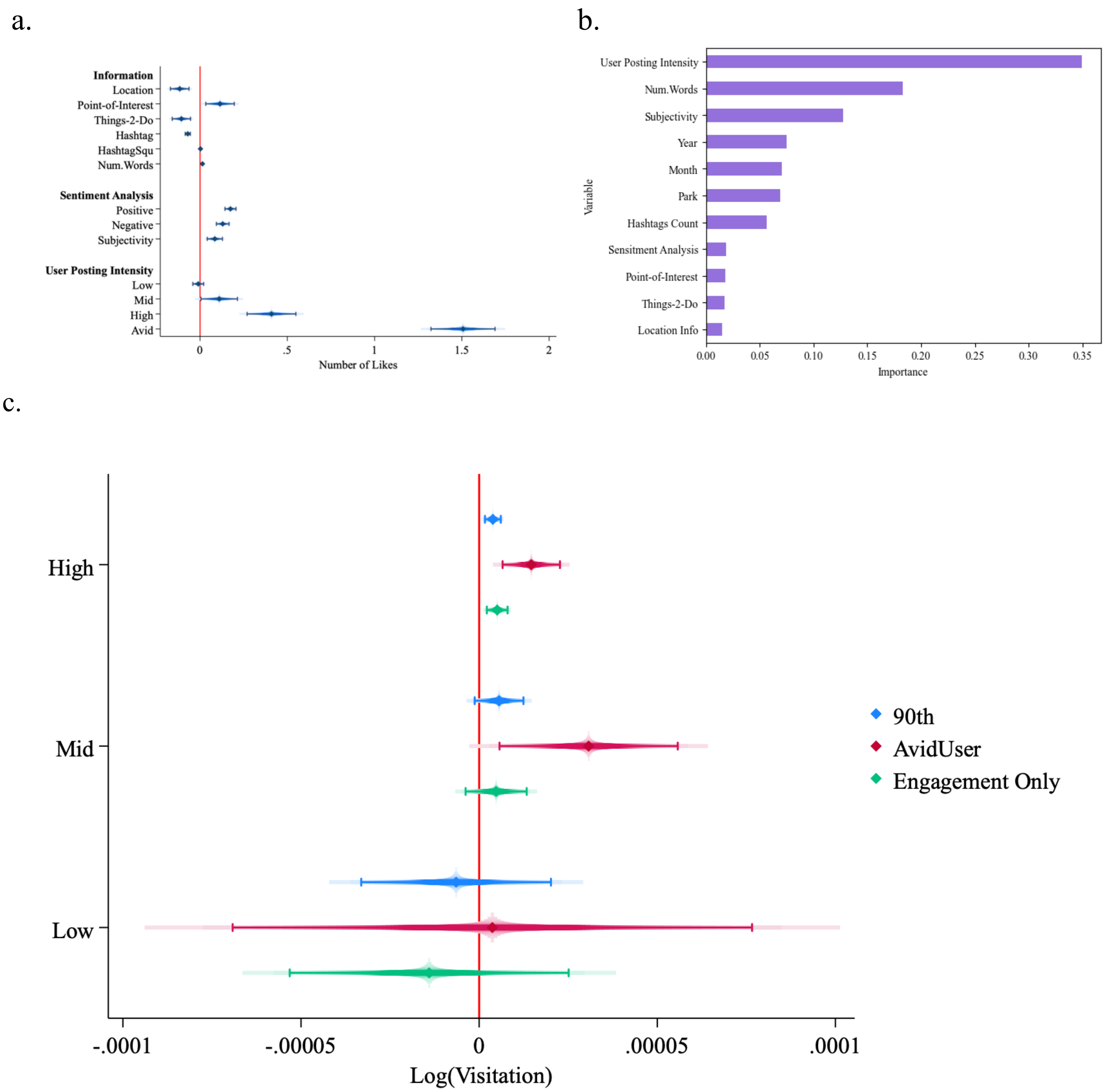
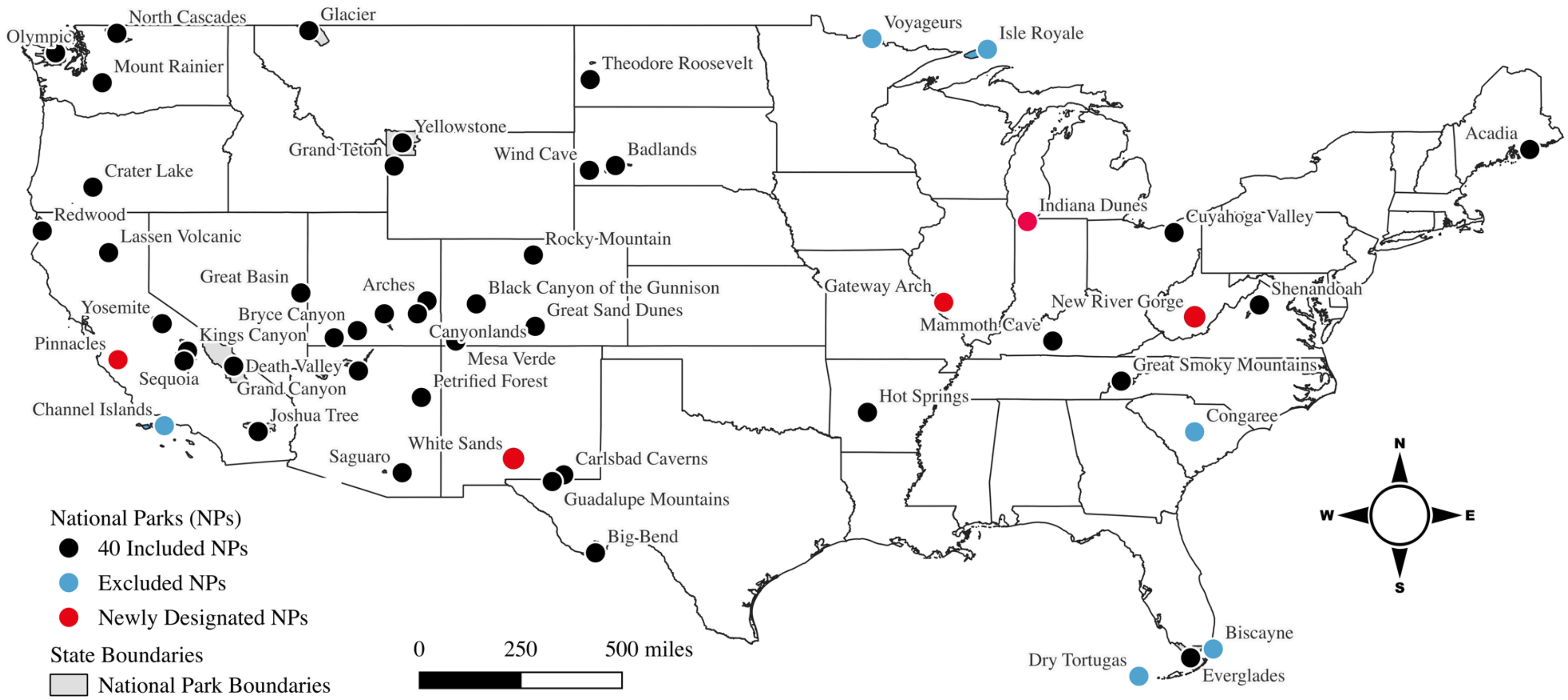


Fig. 6 | Factors determining engagement and the effect of avid users on visitation. Panel (a): The incident rate ratio of the negative binomial count model suggests a 4.5 times higher than all other users. Panel (b): The feature importance from the random forest model suggested user posting intensity has the highest importance rating in the engagement model. Panel (c) reports results from re-estimation of eq. 4 examining the effect of posts in the 90th percentile of highly engaging content (blue), posts from avid users in the highly engaging content category (red) and engaging posts from non-avid users (green). Ticks marks represent the 95 % confidence intervals on estimates.

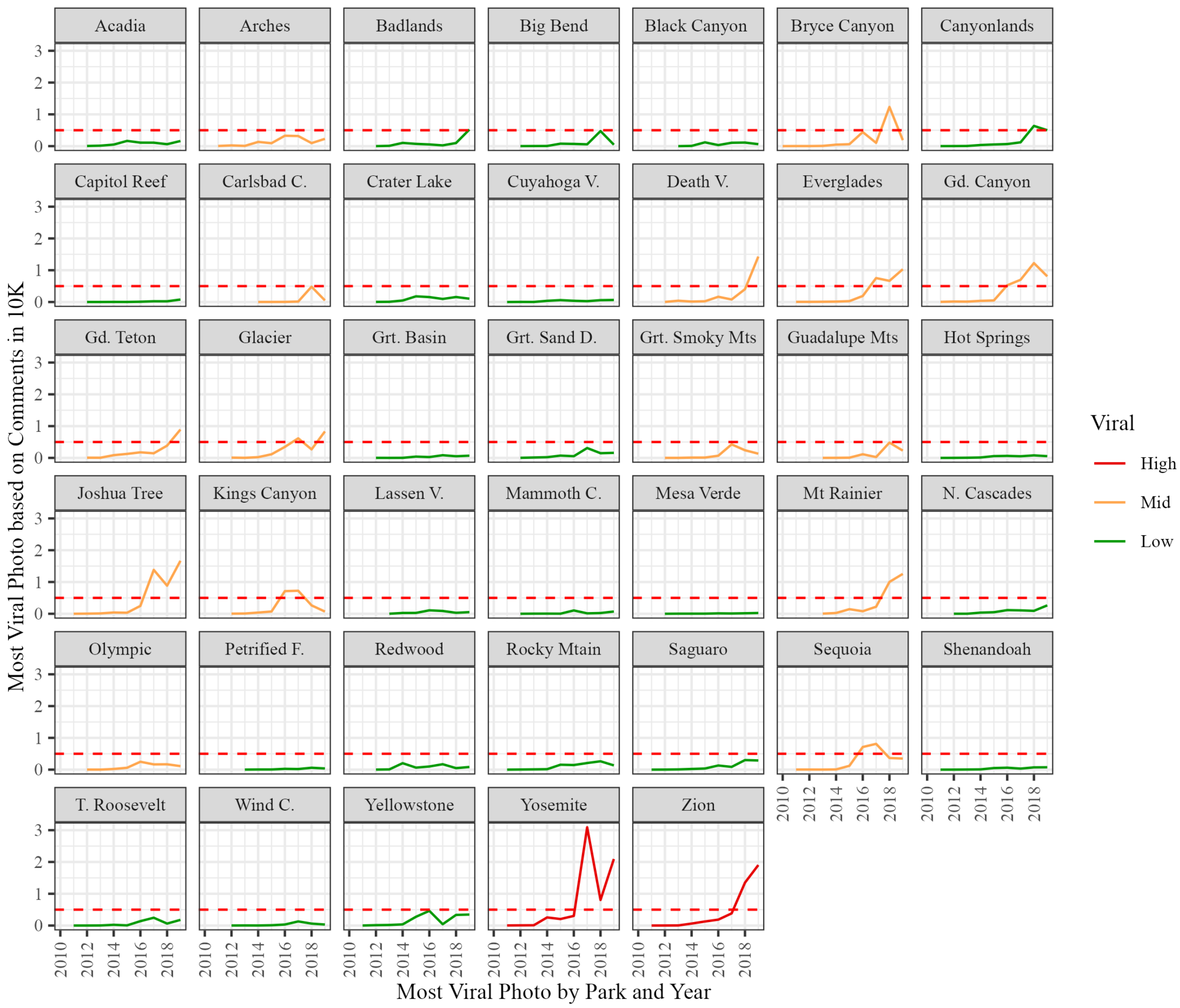
Supplementary Data Figures and Tables

Lowe Mackenzie, A. M. and S. J. Dundas. 2024. *Is a Photo Worth 1,000 Likes? The Influence of Instagram at National Parks*

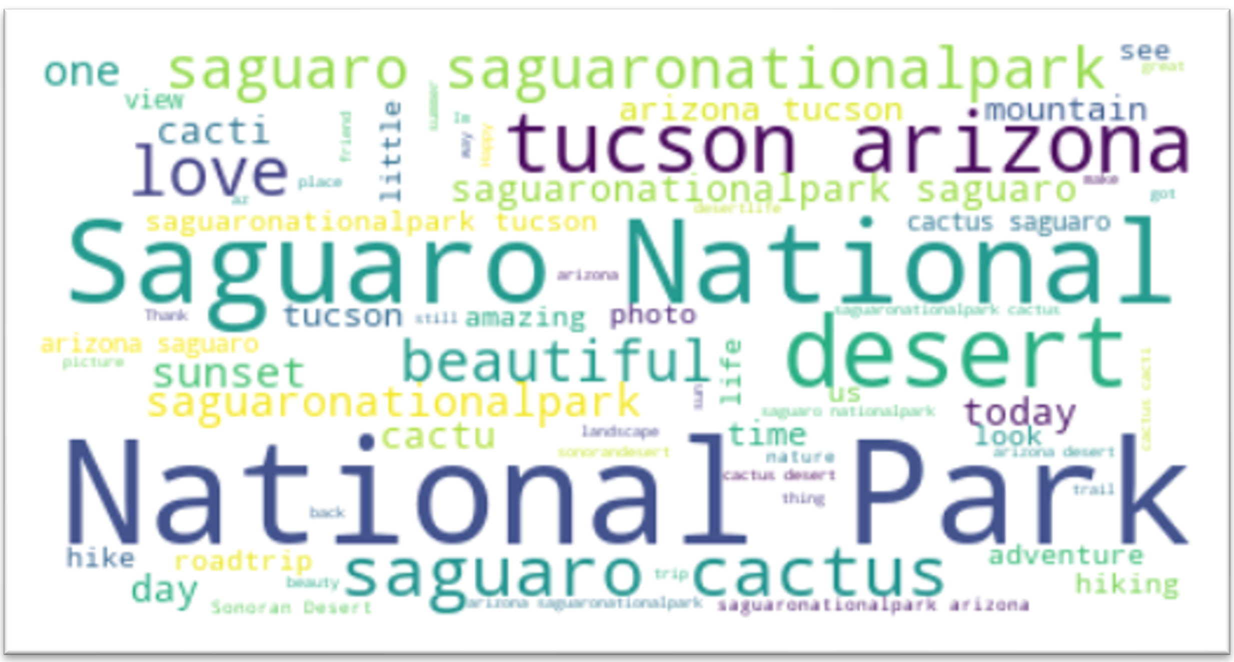


Extended Data Fig. 1: Map of National Parks in the Continental United States of America. Map provides a description of the NPs within the study. The 40 NPs are in black circles. The red and blue circles are not included in the study. Red NPs were classified as NPs post Instagram. Blue NPs are parks excluded due to primary water travel and Congaree due to issues of Instagram retrieval. National Park boundaries are in gray.

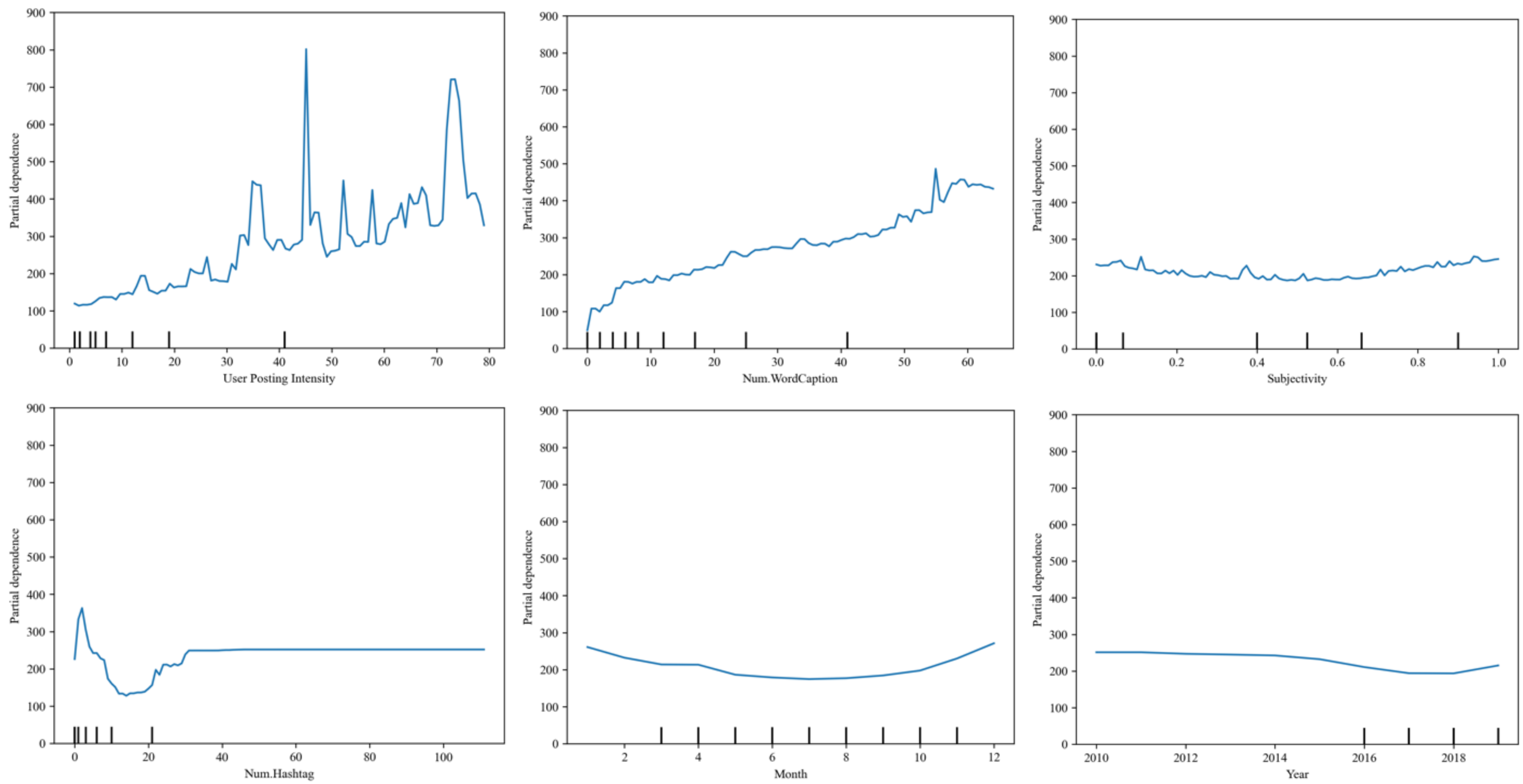
Extended Data Fig. 2: Most Viral Photo Based on Likes for Each Park Per Year. Each park's most viral photo for each year is plotted based on the number of likes the photo received. Parks grouped in low (moderate/high) engagement group are in green (orange/red).



Extended Data Fig. 3: Most Viral Photo Based on Comments for Each Park Per Year. Each park's most viral photo for each year is plotted based on the number of comments the photo received. Parks grouped in low (moderate/high) engagement group are in green (orange/red).



Extended Data Fig. 4: 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. 5: Partial Dependency Plots. The plots are shown for six features of importance from Fig 4.B. and are used to understand the relationship between variable in the random forest 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 Table 1: Descriptive Instagram Statistics for 40 National Parks.

Quantify Set	N	*Likes* to Reach 90th percentile	*Likes* to Reached 95th percentile	*Likes* to Reached 99th percentile
Entire Set of Geotag Photos				
2010-2019	8,310,295	207	347	2,047
Per Year Geotag Photos				
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 2: Textual & Sentiment Analysis w/ Natural Language Toolkit VADER

National Park	Caption (Uncleaned)	Location	POI	Things-2-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
Glacier	▲You don't have to move mountains. Simply fall in love with life. ❤️ Be a tornado 🌪 of happiness ☺️, gratitude 🙏🏾 and acceptance🤷🏾‍♀️. You will change the world 🌎 just by being a warm, kind-hearted human being tbt to this past weekend in montana	1	1	0	Positive
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