

National Park Presentation

INTRODUCTION

National Parks are some of the best places for visitors and residents to explore nature and remember historical events. The United States is lucky to be home to several of these parks. The United States National Parks employ over 20,000 individuals. The amount visitors spend at these parks contributes to the economy. There are over 300 National Parks located throughout the United States, including in U.S. territories of the American Samoa, Puerto Rico and Guam. The National Park Service is primarily funded by Congress, however they also gain funds from park entrance fees and other private philanthropies. For years, the National park Service has lacked the funding it needs to properly protect these parks. The first National Park, created in 1872, was Yellowstone in Wyoming. These National Parks need sufficient funding in order to maintain the beauty of the parks and keep them from deteriorating so future generations can continue to enjoy these beautiful places.

My goal is to analyze the amount of visitation in these parks each year from 1979 to the most recent data, 2021. My analysis focuses on where these parks are located, by state and region and determine which parks contribute the most to the economy and which parks need to increase visitation. I focused on finding correlations between visitation and location as well as economic contribution and visitor spending. I also wanted to look at what impacts COVID-19 had on visitation.

I was able to figure out how visitation, park size, park type, and location correlated to the amount of visitor spending and economic contribution. If I had more time, I would've liked to

gather data on climate change, stocks, and financial patterns to see how these factors could have affected visitation of parks.

DATA

I used data from the National Park Service from 1979 to 2021 to find correlations between visitors, states, parks and regions. I had to scrape data from the National Park Service webpage for these 30 years. I used Python and SQLite to glue the data sets together. Then I brought the data into Rstudio to do more analysis. This data includes visitation for each year and month. This was the primary data I used in my analysis. The *Recreation Visits* variable is the amount of visitors who enter the park, for example, this can be school groups, and tourists. *Non Recreation Visits* would refer to commuters traveling through parks, and NPS employees. I focused on the recreation visits, to show a more realistic count of the visitation at each park. The *Region* variable refers to the subdivisions of the United States as administered by the National Park Services. It includes 7 different regions including Alaska, Intermountain, Midwest, National Capital, Northeast, Pacific West, Southeast and Northeast. The National Capital is part of the Northeast but it is a separate region, only including Washington DC. Another interesting variable was *Park Type*, which included 17 different park types, ranging from National Battlefield, to National Monuments, Parks, and Seashores. Most of these variables were character variables so I cleaned the data by transforming Year and Month variables into integers, and Recreation Visits as numeric. This was a very large dataset, for 30 years of data, I had 173,560 observations and 30 variables. When I started summarizing the data I took a look at the number of parks located in the United States from 1979 to 2021.

I also used All Trails data to see which parks had the most trails and the exact coordinates of these trails and parks. I obtained the All Trails data from Kaggle. All Trails is a website, created for hikers. It acts as a platform to help find the best hiking, running, or biking trails and includes information like reviews, ratings, activities offered, and coordinates of trails. For the All Trails data, I was able to separate latitude and longitude into separate variables in Python so I could use them to make a map visualization. The All Trails data does not include every state, only ones with trails. If I was given more time, I would've liked to look more into the reasons why certain parks are not listed in the dataset. My hypothesis is either the data is not up to date or they only list trails that are listed as hiking trails, rather than trails and walkways that can be found at historic sites, and monuments. While performing summary statistics I was able to find which state had the most trails.

I also included Visitor Spending from 2020 to see how visitation, visitor spending and economic output vary by each state, during COVID-19. The National Park Service provided visitor spending effects, by national economy, state economy, and park economy. However, the Park Data was not readily available to put into R, so I used the visitor spending per state. This data is important in the economic analysis of my data. It shows me which states have the most amount of visitors, how much revenue they got through visitor spending, and how much each state contributed to the economy. This information can also be used to see how the parks help drive tourism and increase outdoor recreation activities. Visitor spending at the parks helps support jobs and benefits the nation's economy. I would have liked to gather more information on the jobs and visitor spending for the 30 year period I looked at to see how it has changed over the years, and if the economic contribution is impacted by any other factors. Given more time I would have also looked at which parks accumulated the most revenue each year.

CONCEPTUAL FRAMEWORK

My hypothesis was that more parks would be located in the West due to less industrialization and more availability of land. During the time when the National Park System was first created, most of the land in the East was privately owned so I hypothesized that parks in the West of the country will generate more economic output and jobs than parks in the East. I also hypothesized that California would be home to the most National Parks, and therefore generate the most revenue and contribute a larger amount to the economy. My model was created to see what correlations can be identified between visitation, location, and economic contribution. If I was able to obtain all the data I wanted, such as climate data and data on the species and wildlife, I would be able to see how visitation was affected by these factors and perform a regression analysis to see what factors played the biggest roles in the amount of spending and economic contribution in the parks. I mostly performed data summaries, to see which regions of the United States are home to the most parks, and what types of parks these are. From there I wanted to tie it to the economic contribution and visitor spending. I made many observations while performing my analysis and hope to continue my research to find out more about the National Parks, and what would be the best way to increase visitation and output in order to gain more funding for these parks. It is known that these parks are some of the best attractions the country has to offer, but they are not correctly funded. In recent news, President Biden has made a plan to make a significant investment of \$3.5 billion in the National Park Service. This money will help strengthen visitor safety, by maintaining the parks. It is meant to help the parks adjust to climate change and reform the infrastructure. This is an important step toward progress in the National Parks. However, I will look more deeply to see how the funding

should be differentiated among the many different parks, and park types. For example, some parks that generate greater revenue might not rely as heavily on this government funding as other less popular parks.

MODELS

For my models and visualizations I used many different layers to form a thorough analysis of the parks. By first determining how many parks are located in the country as a whole, I then branched out to see how many parks are in each region, and each state. Then, I looked to see how many trails each state had in total. I used clustering by using a map I made from the All Trails data to see where parks with the highest amount of trails are clustered. From my initial analysis, I gained a lot of information and began to ponder more questions. A regression analysis can be used to see the relationship between the visitation in these parks and the many different factors that change these rates. Since I was not able to obtain all the data I wanted it was hard to formulate a proper regression. My dependent variable would be the visitation rate, while my independent variables would include the location, type of park, region, climate, amount of trails, and other natural disasters. Instead, I used a lot of clustering techniques to identify different groups within the datasets.

RESULTS

As per results, I found that the amount of parks located in the United States has steadily increased over the past 30 years. In 1979, there were only 268 parks, as of 2021 there are now 368 parks. The number of National Parks has steadily increased over the past 30 years. I looked at which regions had the most parks. I originally thought that the Pacific West Region, which

includes California, would have the most. When I performed my analysis I found that the Intermountain Region which includes Utah, Arizona, Colorado, and Wyoming (just to name a few), had the most parks with 80 as the total. Whereas the Pacific West only had 54 parks, being the fourth most concentrated area of parks. Surprisingly the Northeast and Southeast had the second and third highest count of parks with 72 and 62 respectively. Alaska had the lowest with a total of 15 parks. Western parks have a higher number of National Parks when compared to Northern and Eastern Parks. Alaska has the least number of parks, probably due to the colder climates, and the Midwest region also has very few probably due to the flat land found in that region.

Next I looked at which types of parks were located in each of these regions. Seeing that Alaska had the least amount of parks, it made sense that they also only had 4 different park types, National Reserves being their most significant type. This was interesting, the data showed that the Intermountain region had the most amount of parks, and the region contained a larger variety of park types, it did not have an outstanding park type like the Midwest and Pacific West Regions. For example, the Midwest contains a larger proportion of National Lakeshores and National Rivers. Whereas the Northeast has the most International Historic Sites, and the Pacific West has the most National Reserves. It shows that the Midwest and Southeast regions have a very even distribution of park types.

I wanted to look at the amount of visits per each region and each year. It was clear to see that visitation has steadily increased over the past 30 years, but there is a steep decline in visitation in 2020, during the pandemic. When we look closer, we can see that Alaska has the least amount of recreation visits every year. In contrast, the Southeast and the Pacific West had the most recreation visits on average. When we look at the amount of parks per each state,

California has the most amount of parks, at a total of 24 parks. This is interesting since the Pacific West does not have the most parks in total. But from the visualizations we see that parks in the Northeast, and Intermountain regions, like New York, Pennsylvania, and Maryland have higher amounts of parks, as well as Arizona, and Colorado. We see that Puerto Rico, Guam, and American Samoa all have one National Park. Similarly New Hampshire, Illinois, Connecticut, Rhode Island, and Vermont also have only one.

If we look further into the states that have the most parks, and question which state has the most trails, once again we see that California has the largest number of trails and is an outlier in the data, with 700 trails. Whereas states like Georgia, Missouri, and Michigan look like they have none. In the map visualization of the parks with the most trails, we see the parks with the most trails are clustered in the Pacific West and Intermountain Regions. Some parks that stand out are the Rocky Mountain, Yellowstone, Mount Rainier, the Grand Canyon, and Joshua Tree, all in the Pacific west and Intermountain region. On the East Coast we see the most trails are located in the Great Smoky Mountains, Shenandoah, and Acadia. The map shows that regions with the most parks are located on the coasts rather than in the Midwest Region.

When we look at the visitations by month for the year of 2021, we see a bell curved shape. This shows that more visitors travel to National Parks in the summer months, July being most popular. The winter months have the least amount of visits. When we look by region, Alaska has the lowest amount of visits throughout the whole year, and in the winter months they don't have any visits. Most parks located in the Midwest and Alaska, are visited most during the summer months, while Intermountain parks are visited for a longer duration of the year. Parks with typically warmer climates, located in Southeast and Pacific West regions, vary in visitation

throughout the year but not as much as other regions. In conclusion winter visitation is a lot lower than summer visitation. If I had more data I would've liked to look at the reason for variation in visitation, and find seasonal patterns in each park.

Finally, I looked at the Visitor Spending Data for 2020. Once again California stood out. Having visitation amounting to 28,645,839 visitors, who spent 1,716.5 million dollars resulting in a 2,693.3 million dollar economic output. In contrast, the American Samoa only had 4,819 visitors who spent 0.3 million dollars and contributed to 0.4 million dollars of economic output. We see that Washington DC has a high visitation number but a lot less visitor spending and economic output.

DISCUSSION

The main findings were that California has the most amount of parks and trails, though it is not located in the region that has the most parks. California also has the most amount of visitors per year, so this correlates to why they have the most amount of visitor spending and the highest economic contribution. We can conclude that California does not have to rely heavily on government funding. States with the most parks tend to have higher rates of visitation, and therefore the most visitor spending and economic contribution. I would like to look further into why Washington DC has less economic contribution and visitor spending. My thoughts are that a lot of visits are school bus trips, thus the students are not spending as much as a family would on their visits to the parks. If parks are able to increase their visitation, then they would be able to contribute more to the economy and generate more revenue that could in turn increase the funding that goes toward maintenance of these parks. We saw that the COVID-19 pandemic

affect the parks, by decreasing visitation in all regions. I would like to look further to see if any specific parks increased in visitation during these times.

CONCLUSION

I had many limitations while conducting my research, such as not gathering all the data I wanted and having trouble scrapping the specific data. Gathering and cleaning the data took a lot longer than anticipated and therefore set me back in my research and development of economic analysis. I hope to continue this research and look further into variables that made me question the outcomes. I would like to see how visitation changed for each park over the years, and in each month to see if certain parks and park types have become more or less popular over the years. I would also like to look into the climate changes, and travel patterns to see if these factors have to do with visitation rates. I would look further into the average rating of parks and traffic counts in parks. If I can gather all the variables I need, my analysis would be more thorough and be able to make better suggestions to how the parks can change their current mannerisms to make the National Park System better. For now, we can continue to visit these parks and contribute to the maintenance needed. We can do our part by cleaning up parks while we enjoy hikes, to conserve the beauty displayed at these parks.