***Section 1: Introduction: Domain problem characterization***

The national parks of the United States are known attractions for many travelers and explorers that are seeking new experiences and adventures. For many years, the park authorities have been collecting data that focuses on the visitors to each national park. The problem that we have identified is focused upon the need for greater public awareness as to the utilization level of these parks. By addressing this issue with the development of a data visualization application the public can become aware of the high levels of utilization and work through the proper channels to petition for greater funding. The data set provided covers visitors information from 1904 -2016 and the size of the dataset is 21,560 rows and 18 columns.

The domain of this dataset is tourism/travel. Some of the fields that are within this dataset are the date/year, park name, state, visitors, year, unit name (park name), unit type (park type), and region. We decided to allow the user an opportunity to visualize the location and expanse of the park on a map, provide them with the opportunity to drill down on the raw data of any given park, and to also visualize trends in visitation relative to population growth and gas price fluctuations. This display approach will help make actionable visualizations for this large dataset.

***Section 2: Data/Operation abstraction design***

The National Park data set was found on dataworld.com ([https://data.world/inform8n/us-national-parks-visitation-1904-2016-with-boundaries](about:blank)). The data sets were extracted as CSV files. The first and also largest data set was the park visits data set. This file offered the park name, region and number of visitors. The next data set was the state population data set featuring the year, state, and population of the state. The third piece was the gas price data set which featured the year, current gas prices for the year, and constant gas prices for the year. The latitude and longitude features were sourced from an online embedded table an entered manually into a .csv file. Read in as a character data type, they were converted to numerical values so that they could be properly used within the app.

To tidy the data, the datasets were joined. The first data sets that were joined were the park visits and gas price data sets. A left join was performed on these data sets. Next, this data set was joined with the state population file. A left join was performed for these sets, as well. Last, this data was combined with the locations data set, which had been uploaded to the console.

The Park Visits Year over Year data set was created once all the data sets had been joined. Year and visitors were the two features that were selected from this data set. Year “total” was filtered out from the data set, as well has years that were less than 1950, to allow for a more concise view of the trends across parks. The ‘Year’ column was mutated to an integer so it could be properly used within the app. The data set was grouped by and arranged by year. A new column was created “visitors\_total” which is equal to the sum of visitors. The State Population Year over Year data set was created by using the same exact code, but instead of selecting “year” and “visitors”, you select “year” and “pop,” Also, for the summarise function we utilized the output was titled “pop\_total” which equals the sum of the population of all states. Park visits year over year was joined with state pop year over year. This data set was then joined with the gas price data set. Both joins were left joins.

Once this data was aggregated in a single file, we utilized it to create a second file which provided year over year trend data, both in absolute terms and on a percentage basis. This data underlines the line charts featured in the app.

For the map of the USA within the app, we utilized the rworldmap package. We then layered in the location of each park, which was read in from a file manually constructed by the team. Any unit types that did not equal “National Park” were filtered out. Years that were less than 1950 were filtered out as well as well to focus the user on more relatable time periods.

***Section 3: Encoding/Interaction Design***

The following libraries were used for the design of the app:

* shinyWidgets
* dplyr
* rworldmap
* ggplot2
* tidyverse
* rsconnect
* shyndashobard
* readr

For the design of the app, we focused on providing the user multiple ways to access the data. We provided a map as the first output of the app to allow the user to visualize the location of the parks and the relative attendance amongst them. We believe that this familiarize the user with the park system, the overall size of it, the general expanse of the system, and personalizes the system as it helps the user think of which parks they have visited or perhaps which areas of the country offer national parks so that on their next visit they may choose to visit. We wanted the parks to be thought of as part of the user’s life, not an abstract resource. We furthered this goal by including a gif of Smokey the Bear stating ‘Think’ on the next page of the app. Smokey the Bear is a familiar character to most Americans and brings continuity to park conversation efforts which have spanned decades.

The next feature of the app is a simple table output. Our hope in placing the table output at this point in the app was to allow the user to pick their favorite park and spend a moment with the data. This will hopefully have the result of making the user realize that the national parks our the responsibility of all. Rather than thinking of the parks as a system, the user now thinks of a specific park where they may have a cherished memory of a family outing, hiking experience, or vacation. We want the user to identify with the park system through their personal favorite. Now that the user is thinking about how they park system has impacted their own lives, we behoove them on the next tab to ‘Take Action’, again using Smokey the Bear. This time Smokey is with a troop of boy and girl scouts. This aims to remind the user that the parks our also a resource for future generations.

##### The next tab features trend data around visitations, population growth, and gas prices. The charts are constructed to allow the user to add data series through selection. However, the main takeaway is the increased usage of the parks by the citizens of the country. This is emphasized in two ways. First, we created a variable which reflects the ratio of country population to park visitations. We impose this on the same graph as gas prices and population trends. To make sure the point is not missed, we included an explanation box which specifically mentions the variable ‘visitpopper’, which we designed to specifically highlight this trend we noticed in the data while wrangling it.

##### At the bottom of this tab, we again emphasize this trend by creating a graph of the number of visitors and the country population. We defaulted the years on the slider to show a clear increase of visitations which began in the 1950s. We show that ratio of country population to park visits has grown to almost 1:1. It would be difficult for the user not to have awareness of the extent that our parks are being utilized by the citizenry.

##### The last tab features a beautiful gif of Half Dome in Yosemite National Park. It is one of the treasures of our park system and is also featured on the opening tab, with a long line of hikers. This last image is meant to inspire the user and features a call to action. Let’s be part of the solution! For purposes of presentation, we then can scroll down to a celebratory dancing bear for our Q&A backdrop. However, for dissemination purposes, we would not include this last gif, it is only for our class presentation.

Overall, our design intent was to tell a story regarding our national parks and invoke an emotional reaction from our user as to spur them to action. We used a mix of data and visualizations to accomplish this task.

1. ***Algorithmic design***

The dataset offered by the National Park Service is very dense. Our algorithmic goal was to utilize only the data we needed. As part of our tidying, we removed much of the data from the main file and created smaller, more agile data files which only contained the information necessary for that portion of the app. This helped aid the execution speed of the app.

We also decided to utilize a simple approach on the map. Though the park data included shape files that would have allowed for the mapping of the actual park contours onto the map, we didn’t feel that information was necessary for our goal. We simply wanted to bring awareness to expanse of our national park system, not necessarily the shape or size of any individual park, so we simply used a plotting approach through latitude and longitude data.

1. ***User evaluation***

We believe that we have created a fun app! The National Park Service is a respected entity in our cultural yore, even if it only reminds the user of Yogi Bear and Jellystone National Park! If implementing this awareness campaign more broadly, we believe that this app can easily be shared via a social media campaign that would offer users the opportunity to explore the data while the integration of the fun gifs will provide some level of entertainment for the user. The utilization of the app would give us a strong indication of the effectiveness of the campaign. In addition, depending on the target audience different versions of the app could be created that feature different pictures and gifs. For example, for younger target audiences we might utilize SpongeBob rather than Smokey Bear or Yogi Bear. For campaigns targeting a specific geographic location, we can utilize pictures of a nearby park. For example, if we were advertising the campaign on the website of Southern Living, we would make sure the visuals were focused on Great Smoky Mountains National Park. If we were advertising the campaign on the website of Boston Magazine, we would include pictures of Acadia National Park. The core of the app would be leverageable across all markets, but swapping the art work would make it feel customized to the audience.

***6. Future work***

We have three main areas of improvement that we would focus upon if further enhancing our application. The first would be the utilization of the plotly package to allow for a more interactive user experience. While we had originally used the plotly package in designing our map and charts, we had difficulty in getting the plotly package to execute within the shiny app, though the code worked fine in a plain R script environment. We had to remove the plotly language from the app to allow us to move forward. Our next version of the app would include work around making the plotly outputs viable within the Shiny environment.

Secondly, we would have liked to enhance the look of our map. The googleVis package is really neat and allows for striking visuals from your graphs. Given the natural beauty of our national parks, a map which featured terrain aspects of the United States would have added to the opening impact that our map is designed to induce. The googleVis packaged now requires registration and verification code for Google. The work to register with google and integrate the googleVis package into our app would be worthwhile in our next iteration.

Lastly, we had difficulty getting our app to publish on shinyapps.io. Our next iteration would need to do so as to support our social media campaign, so we would need to work with support at RStudio to trouble shoot the issue inhibiting the publication.

***Links***

<https://rstudio.cloud/project/775599>

<https://github.com/bdignan75/UNCC5122-Final-Project>