Outdoor Equity App Technical Documentation

Clarissa Boyajian and Halina Do-Linh

2022-06-02

Contents

Chapter 1

About

1.1 Abstract

Outdoor recreation and access to nature have well-documented positive impacts on mental and physical well-being. Federal public land management agencies in the United States offer a variety of outdoor recreation activities to visitors. However, people from different socioeconomic and identity groups access federal public lands unequally due to historical discrimination and current inequities. This project uses data from the Recreation Information Database (RIDB) and the United States Census Bureau (US Census) to explore patterns of visitor use of reservable overnight sites (such as campgrounds, cabins, hike-in, and more). Specifically, we used 2018 reservation data and US Census data from the next available year to 2018 (i.e. 2018 median income data, 2015 language data). We created the interactive Outdoor Equity App that gives users tools to summarize data, explore relationships between RIDB and US Census variables, view maps of where visitors are coming from for reservable sites in California, and download subset data. This technical documentation includes information on metadata, application maintenance, and next steps for expanding the app to include visitor data from more locations and time periods.

1.2 About the Authors

This technical documentation for the Outdoor Equity App was created by Clarissa Boyajian and Halina Do-Linh. The app was created as the final capstone project for their Master of Environmental Data Science degrees from the University of California's Bren School of Environmental Science & Management. Both women are passionate about environmental justice, open science, the art of data visualizations, and spending time recreating outdoors. Please reach out to either of both of us with any questions.

This project could not have been completed without the support and guidance of the Bren School advisors Dr. Frank Davis and Dr. Allison Horst and our external advisors Dr. Kaitlyn Gaynor and Dr. Will Rice.

1.3 Helpful Links and Resources

The Outdoor Equity App is was created with the `shiny` package [Chang et al., 2021a] using RStudio version 1.4.1717-3. This technical documentation is hosted using GitHub Pages. The GitHub repository containing all code relating to this technical documentation can be found here and the GitHub repository containing all code relating to Outdoor Equity App can be found here.

Chapter 2

Executive Summary

Outdoor recreation and access to nature have well-documented positive impacts on mental physical well-being. Federal public land management agencies in the United States offer a wide variety of activities to visitors. However, people from different socioeconomic and identity groups access federal public lands unequally due to historical discrimination and current inequities. The multi-agency program, Recreation One Stop (R1S), oversees the operations of Recreation.gov and aims to increase access to recreation by providing online resources about nationwide recreational opportunities, allowing visitors to make reservations, and making the associated data accessible to all. The rich data on visitors that R1S collects presents an opportunity for the creation of more robust data-driven analytical tools to understand the patterns and correlations of this unequal access across the country and within individual recreation areas. Decision-makers can use these tools to explore and visualize how recreational opportunities on federal public lands are accessed.

Our overarching objective is to design and built an interactive web application that allows users to analyze patterns in the access and demand of visitors at reservable overnight sites (such as campgrounds, cabins, hike-in, and more), using data from the Recreation Information Database (RIDB) and the United States Census Bureau (US Census). These analyses will allow federal public land managers to explore relationships among attributes of recreation opportunities, reservation practices, and socioeconomic data from the regions of visitor origin. We achieved this goal through the creation of an interactive web application, the Outdoor Equity App, that allows for a wide range of visualization, metadata documentation, and subset data downloads. This technical documentation serves to document the Outdoor Equity App creation process, include information for ongoing maintenance, and provide suggestions for future use and expansion.

The app - which is implemented using the R programming language - accesses public RIDB and US Census data via direct download and application programming interfaces (API). All data and R code scripts are stored on the UCSB Taylor Server and version-controlled through GitHub. We isolated necessary variables and defined, standardized, and aggregated values in the data cleaning process. We calculated additional derived variables for each reservation, such as distance traveled and booking window, and summary statistics (e.g., mean and median) for census data at the ZIP code level. A data set that combines the US Census and RIDB data based on visitors' home ZIP code is the foundation for the Outdoor Equity App. We visualized distributions of variables and relationships between them with simple, straightforward figures. Within the app, users can subset the data to a specific overnight reservable site and visualize the distribution of a single variable, the relationship between two variables, or the visitorshed map (i.e. area from where visitors are coming) for the selected site. The app currently only includes data for California reservable sites in fiscal year 2018 due to project scope limitations.

Throughout the analysis and app creation processes, external advisors and federal public land managers have reviewed and tested the Outdoor Equity App. We incorporated feedback into all parts of the processes to ensure our data, analysis, and final products are robust. Potential future updates to the Outdoor Equity App are discussed in this technical documentation and include temporal and spatial expansions and app maintenance. The temporal expansion would include cleaning additional datasets for years from 2012 to 2021 as well as expanding the app's interface to allow for temporal selections when subsetting data. The spatial expansion would focus on updating the app structure and server hosting capabilities so the app runs smoothly with data from the full United States.

As environmental justice is increasingly recognized as a necessary lens to achieve environmental goals, equitable access to outdoor recreation is a high priority for managers. This tool assists managers to be equity-conscious decision-makers, can be a springboard for researchers who have questions about outdoor recreation, and strengthens nonprofit organizations' advocacy efforts. We also hope it will be a dynamic tool that empowers visitors to access the information and resources they need to explore outdoor recreation.

Chapter 3

Problem Statement

3.1 Background

Outdoor recreation provides critical health and well-being benefits to communities, and in the United States, federal public lands play an important role in providing access to nature. However, access is not equal for all people [Ewert and Hollenhorst, 1990]; [Flores et al., 2018], which has been recognized as an environmental injustice [Floyd and Johnson, 2002]. Many studies have shown that federally managed public land is accessed unequally due to historical discrimination and current inequities [Floyd and Johnson, 2002]; [Shelby et al., 1989]; [Xiao et al., 2021].

Many land management agencies in the U.S. are tasked with the dual mandate of providing recreational opportunities for visitors while also preserving and conserving natural resources and places [Shartaj and Suter, 2020]. For over a century, striking the balance necessary to uphold this mandate has proven a challenge for federal agencies like the National Parks Service [Meinecke, 1937]; [Sax, 1980], and the recent growth of recreation (Figure 3.1) has renewed concerns about its potential negative environmental impacts and changes to the visitor experience [Hammitt et al., 2015]; [Timmons, 2019]. The challenge now facing public land management agencies is how to allocate quality visitor experiences to a more diverse user base. Simply increasing recreation opportunities on public land is not a viable solution to this rising demand.

While managers seek to allocate existing resources (e.g. campsites) through the fairest means possible, including reservation systems, equal opportunities do not translate to equitable access [Shelby et al., 1989]. Historically, policies of segregation barred certain racial groups from using federal public lands and the legacy of these policies has perpetuated inequitable access for certain racial groups to this day [Xiao et al., 2021]. Additionally, previous and current inequities like lack of time, disposable income, access to technology, and lack of social or institutional knowledge about reservation systems impact access to federal public lands [Scott and Lee, 2018]. At present, park visitation and camping are seeing a surge in popularity, heightened even more by the COVID-19 pandemic, and this rapid increase in demand for recreation opportunities may only further these inequities.

3.2 Significance

Currently, much of our understanding about trends in recreation on public lands comes from the Integrated Resource Management Applications (IRMA) Portal, which the National Parks Service uses to monitor visitor counts over time [Bergstrom et al., 2020]. However, these data lack information on where visitors are coming from. This project leverages the Recreation Information Database (RIDB), managed by Recreation One Stop, an inter-agency partnership that provides reservation services and trip-planning tools on Recreation.gov. The RIDB is far more robust, including data from other land management agencies, and information on visitor zip codes, costs, group sizes, and dates of both reservations and recreation activities. While it is available for public download, there are few robust data-driven analytical tools to understand the patterns and relationships of these inequities within individual recreation areas.

Previous research has demonstrated the value of RIDB data in forecasting future recreation demand for single park units [Rice et al., 2019] and analyzing preferential characteristics for popular recreational facilities [Rice and Park, 2021]. A recent study summarizing RIDB data from national parks [Walls et al., 2018] also identified broad patterns in reservations. For example, campsite reservations are made far in advance, but many are canceled last minute (Figure 3.2); visitors tend to visit national parks near their homes (Figure 3.3); and the distribution of incomes of campers appears to be similar to the U.S. population as a whole (Figure 3.4). However, overall, the vast RIDB data has received limited system-level research attention to date, and this work will be the first to explore issues of equity with RIDB data.

Furthermore, much of the existing research on outdoor recreation focuses on National Park Service lands, such as Walls et al. [2018], which is only a small percentage of all federal land used by the public. The other land management agencies, including US Forest Service, Bureau of Land Management, and Army Corps of Engineers, often lack the capacity and funding to process reservation data, and are less frequently the subjects of outside research. Little is known about how patterns of access and demand vary across land management types. The RIDB includes data from all federal land management agencies, and therefore has tremendous promise to inform our understanding of patterns and trends in recreation across space and time and to inform policies for more equitable campground access for all federal public lands.

Our overarching objective for this project is to utilize data from RIDB and US

Census to analyze spatial and demand patterns of visitor access at reservable overnight sites (such as campgrounds, cabins, hike-in, and more). We chose to focus on reservable sites since recent studies have shown this type of outdoor recreation to be a good proxy for visitation to federal public lands [Walls et al., 2018]. These analyses will provide federal public managers an opportunity to explore relationships between and within socioeconomic and reservation variables.

3.3 Figures

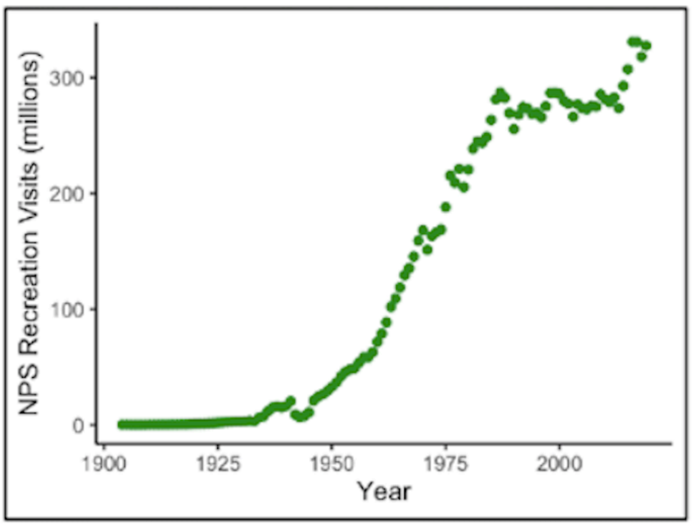

Figure 3.1: Total annual visitors to the National Park Service system, since its inception through 2020. Visitation has been rapidly increasing, particularly within the last decade. (Source: [IRMA](https://irma.nps.gov/Portal/))

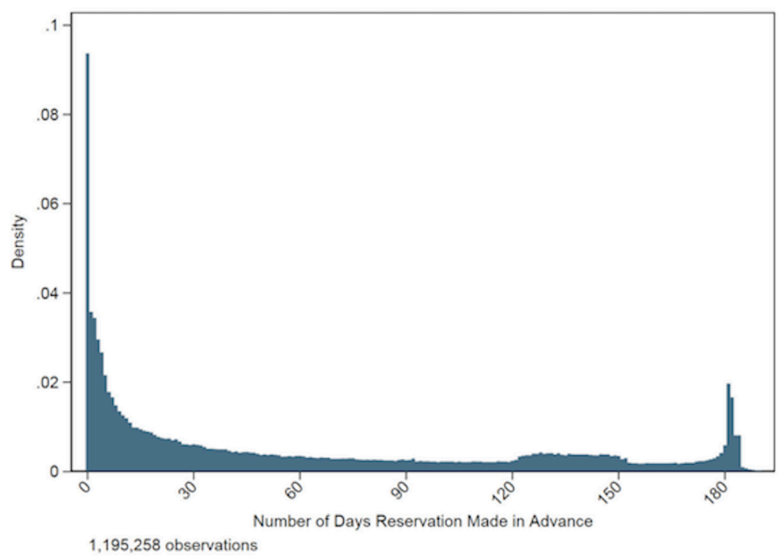

Figure 3.2: Reproduced from @Walls2018. Days in advance that National Park campsite reservations are made from 2014 to 2016. Reservations are made far in advance, but many reservations are canceled at the last minute.

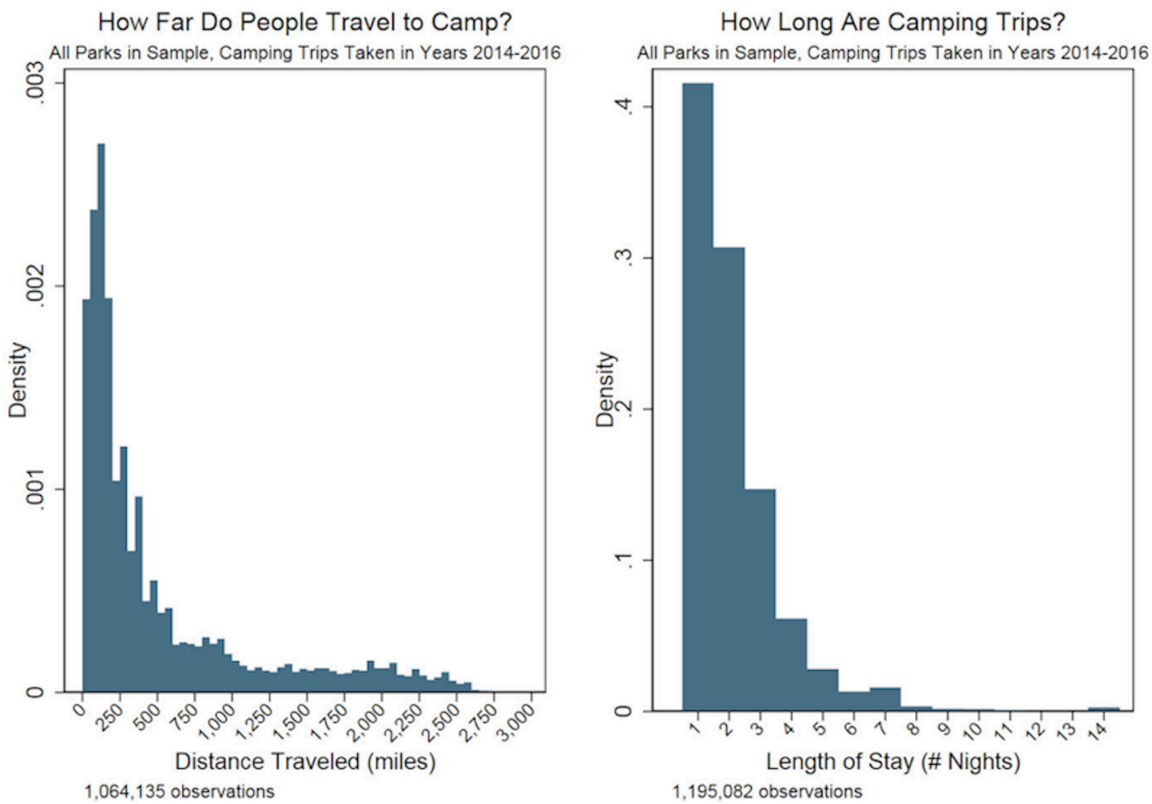

Figure 3.3: Reproduced from @Walls2018. Distance traveled and duration of stay for National Park camping visits from 2014 to 2016. Visitors tend to visit national parks near their homes and stay only two nights, and longer trips are rare.

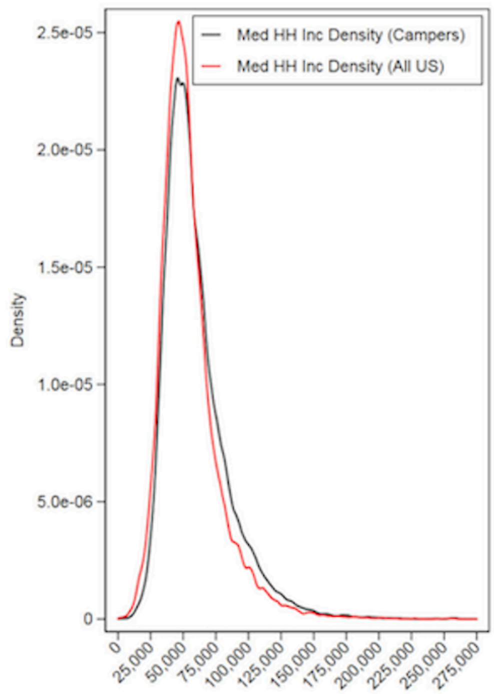

Figure 3.4: Reproduced from @Walls2018. Median household income (by zip code) for National Park campers and for US Population. The black line estimates the distribution of median household (HH) income of campers from 2014 to 2016. The red line estimates the distribution of median household for all zip codes in the U.S. using average median household income from 2014 to 2016 where each zip code is an observation.

Chapter 4

Specific Objectives

Federal lands in the United States provide important recreation opportunities to the public, but there is a **growing need to understand and mitigate inequities in access to outdoor recreation.** This project addressed this need by creating the Outdoor Equity App, an **interactive platform** for summarizing and visualizing site-specific patterns and trends in visitation volume, demand, and visitors' location of origin. The platform will integrate nationwide Recreation.gov reservation data with US census data to:

- Gain insights into **demand for reservations** across different types of recreation areas.
- Analyze access to federal public lands among **historically underserved groups** in relation to recreation **site type, cost, location, and demand**.
- Clearly define all variables and values in the **metadata documentation**.
- Allow users to **download a subset of the combined data** for further analysis.

Chapter 5

Summary of Solution Design

5.1 Glossary and Definitions

Throughout this document we define "reservable sites" as traditional campgrounds, single remote campsites, overnight boat-in sites or mooring, equestrian sites, cabins, and other shelters listed in the RIDB data.

5.2 Access, Clean, and Wrangle Data

RIDB data and US Census American Communities Survey (ACS) data are freely available online to the public. We accessed, cleaned, and wrangled all data

Table 5.1: A table of abbreviations, their definitions, and source URLS.

Abbreviation	Definition	Source
ACS	American Community Survey	https://www.census.gov/programs-surveys/acs
BLM	Bureau of Land Management	https://www.blm.gov/
BOR	Bureau of Reclamation	https://www.usbr.gov/
MEDS	Master of Environmental Data Science	https://bren.ucsb.edu/masters-programs/master-environ
NPS	National Park Service	https://www.nps.gov/index.htm
R1S	Recreation One Stop	https://www.recreation.gov/
UCSB	University of California, Santa Barbara	https://www.ucsb.edu/
USACE	United States Army Corps of Engineers	https://www.usace.army.mil/
USFS	United States Forest Service	https://www.fs.usda.gov/

Table 5.2: A table of the cleaning and wrangling functions created for the ACS and RIDB data and functions to create data sets for visitorshed maps and data relationship plots.

Script Location	Purpose
data_preparation/functions/function_acs_deciles_median_income.R	Calculate decile values of
data_preparation/functions/function_acs_education.R	Call and calculate educat
data_preparation/functions/function_acs_language.R	Call and calculate langua
data_preparation/functions/function_acs_median_income.R	Call and calculate mediar
data_preparation/functions/function_acs_race.R	Call and calculate race pe
data_preparation/functions/function_acs_top_quartile_education.R	Calculate weighted third
data_preparation/functions/function_acs_top_quartile_language.R	Calculate weighted third
data_preparation/functions/function_acs_top_quartile_race.R	Calculate weighted third
data_preparation/functions/function_ridb_subset-pre2018.R	Subset RIDB data
data_preparation/functions/function_ridb_variable_calculate-pre2018.R	Define, standardize, and a
data_preparation/functions/function_join_ridb_acs.R	Join RIDB and ACS data
data_preparation/functions/function_map_ca_data.R	Create dataset for Califor
data_preparation/functions/function_map_us_data.R	Create dataset for US Sta
data_preparation/functions/function_ridb_deciles_median_income.R	Create dataset for mediar
data_preparation/functions/function_ridb_top_quartile_education.R	Create dataset for educat
data_preparation/functions/function_ridb_top_quartile_language.R	Create dataset for langua
data_preparation/functions/function_ridb_top_quartile_race.R	Create dataset for race da

outside of the Outdoor Equity App using the `data_wrangle_and_clean.Rmd` document and 18 custom-made functions (Tables 5.2). We downloaded RIDB data in CSV format from Recreation.gov and ACS data through API using the R package `tidycensus` [Walker and Herman, 2022]. We first subsetted RIDB and ACS datasets to include only the variables relevant to our objectives. We then normalized, aggregated, and calculated variables as necessary. Once both datasets are cleaned and wrangled, we joined them using ZIP codes as the key (common value in both datasets). Finally we wrangled the joined RIDB and ACS dataset to ready them for creating data relationship plots.

5.3 Analysis and Visualizations

The Outdoor Equity App features interactive maps and plots. Users of this app can select a single reservable site to create custom plots that show a data summary of a single variable or a data relationship between two variables. Visualizations of multiple reservable sites appear as separate plots. Users can also select a single site to create a visitorshed map for the full United States and for the state in which the site is located.

5.4 Outdoor Equity App

The app has a navigation bar with four tabs: About, Analysis, Metadata and Data Download. Nested under the Analysis tab are the subtabs of Data Summary, Data Relationship, and Visitorshed Maps. The app opens automatically to the About tab.

Figure 5.1: Screenshot of the About page of the Outdoor Equity App

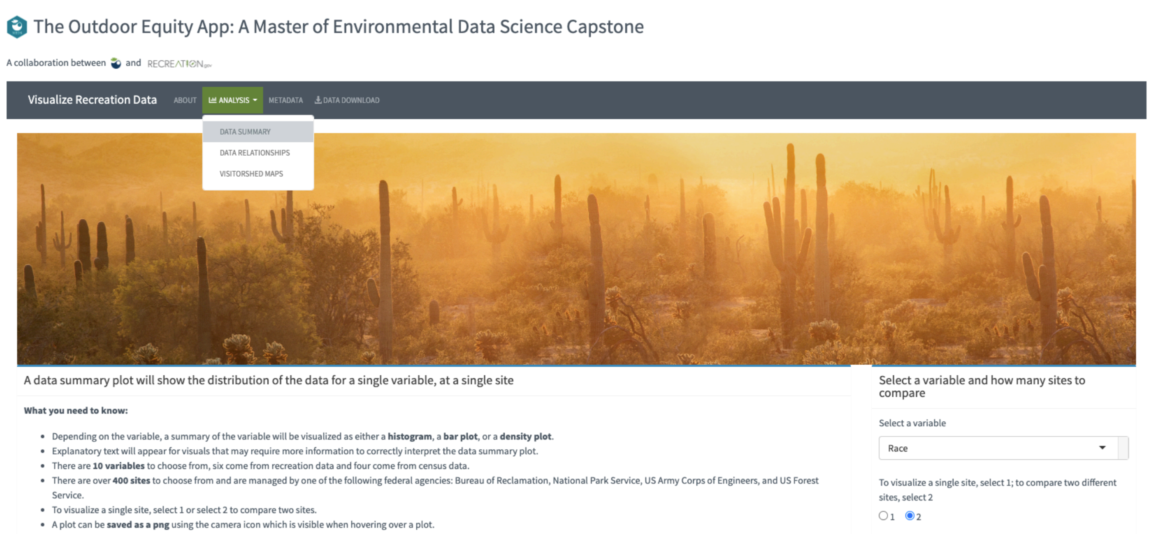

Figure 5.2: Screenshot of the Analysis page of the Outdoor Equity App

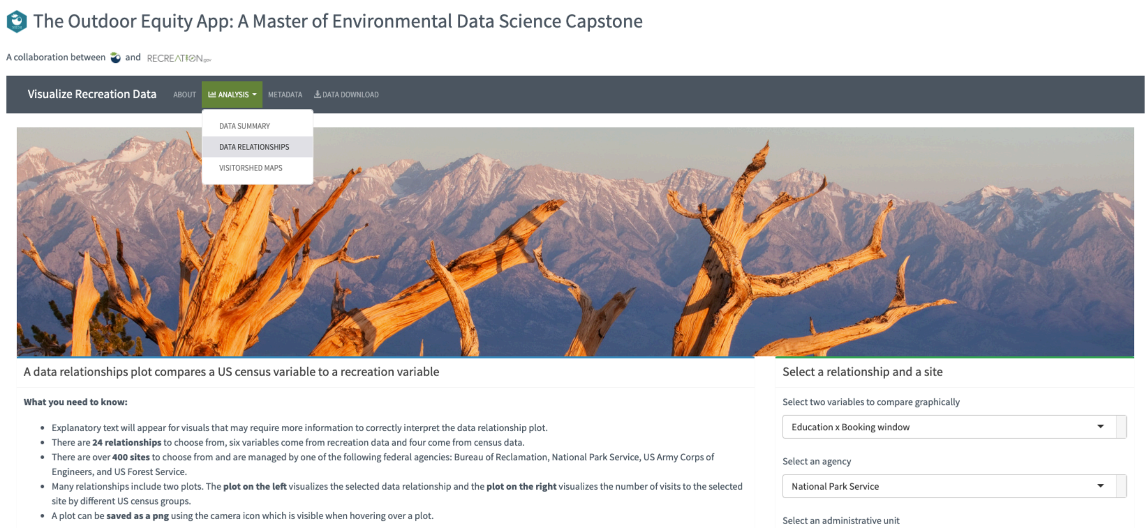

Figure 5.3: Screenshot of the Analysis page of the Outdoor Equity App

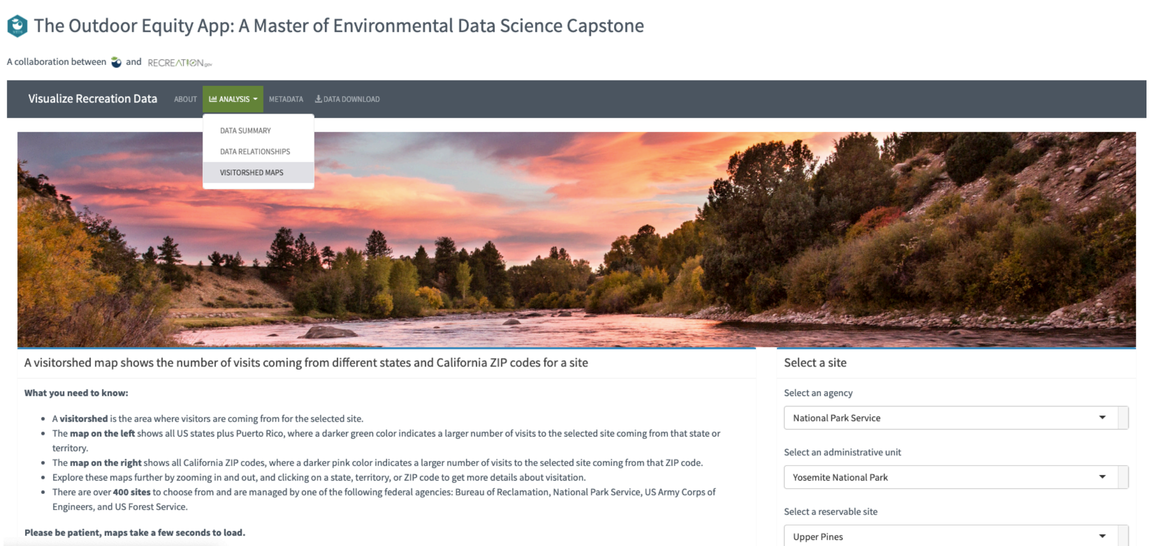

Figure 5.4: Screenshot of the Analysis page of the Outdoor Equity App

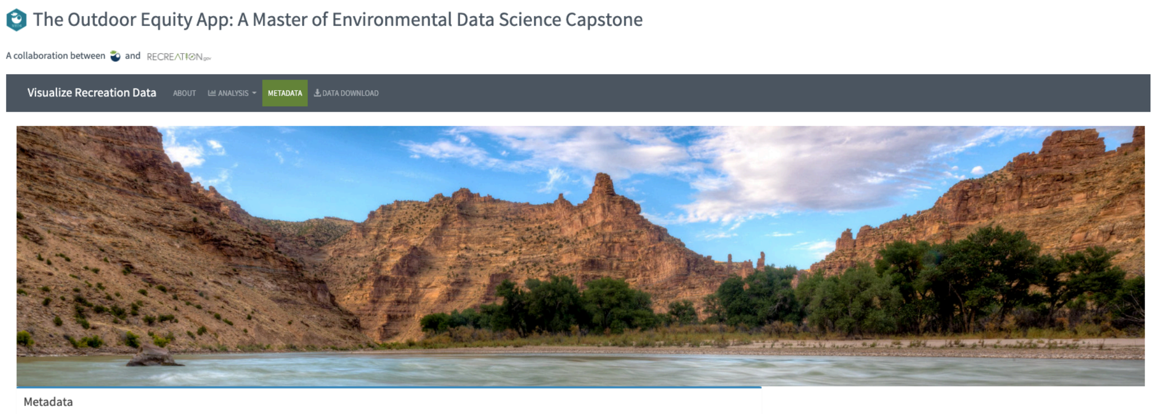

Figure 5.5: Screenshot of the Metadata page of the Outdoor Equity App

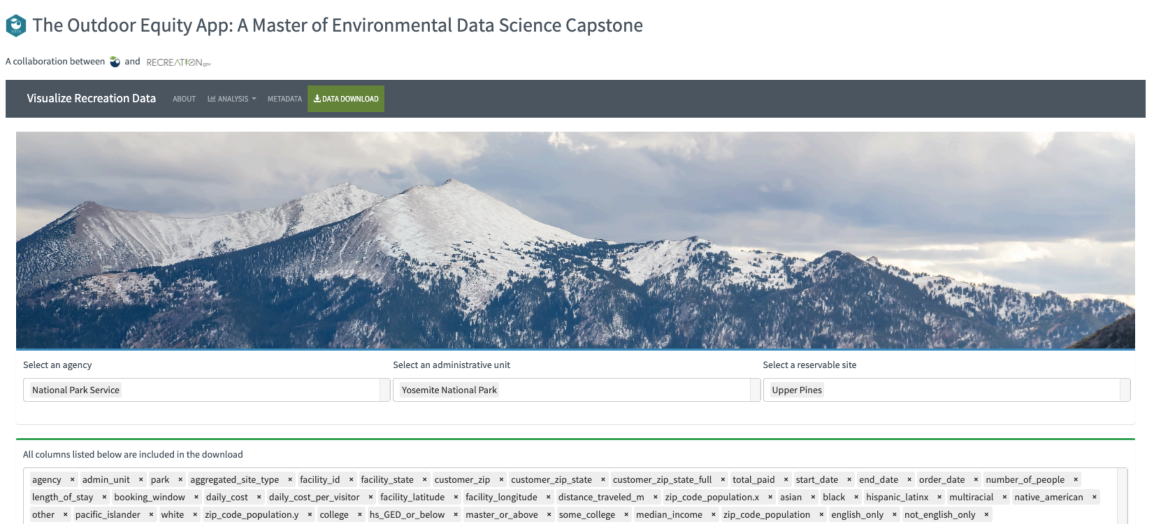

Figure 5.6: Screenshot of the Metadata page of the Outdoor Equity App

5.5 User Guide and Technical Documentation

The Outdoor Equity App includes a user guide and metadata information. The user guide section includes a quick overview of the app and helper text on how to start creating visuals.

This technical documentation is created with the `bookdown` package [Xie, 2021] and is linked in the About tab of the Outdoor Equity App. Metadata for all variables used within the app are also available on the app Metadata tab and in the Products and Deliverables Section of this document.

Chapter 6

Products and Deliverables

6.1 R Shiny App

The Outdoor Equity App was built using the `shiny` package [Chang et al., 2021a] and has the following functionality:

- Visualize statistical distributions of RIDB and US Census American Community Survey (ACS) variables
- Visualize relationships between RIDB and ACS variables
- Download visualizations as PNG images
- View visitorshed maps of reservable sites both nationally and within the state where the campsite is located
- Download customized subsets of the data

6.2 Metadata

6.3 Package Versions

The table below includes information on the specific package versions used to create the Outdoor Equity App.

Table 6.1: Metadata of joined RIDB-ACS dataset

Variable Name	Data Source	Definition
agency	ridb	the governing body that manages a type of US public land
admin_unit	ridb	the parent location or region description that a campsite b
park	ridb	the name of a campsite
aggregated_site_type	ridb	type of site at a campsite; a campsite can have multiple sit
facility_id	ridb	unique id given to a campsite
facility_state	ridb	the state that a campsite is located in
customer_zip	ridb	the numeric code of the area from where a visitor lives
customer_zip_state	acs	state acronym for home state of visitor
customer_zip_state_full	acs	full name of state for home state of visitor
total_paid	ridb	total amount of dollars paid for a reservation
start_date	ridb	date when booked reservation begins
end_date	ridb	date when booked reservation ends
order_date	ridb	date when reservation was booked and purchased
number_of_people	ridb	number of people reported when booking reservation
length_of_stay	ridb	the number of days a visit is; difference of end date from st
booking_window	ridb	the number of days a reservation is made before the start o
daily_cost	ridb	the total amount paid per day for a reservation
daily_cost_per_visitor	ridb	the total amount paid per day for one person
facility_latitude	ridb	latitude of the campsite, but note this may not be the cent
facility_longitude	ridb	longitude of the campsite, but note this may not be the cen
distance_traveled_m	ridb	distance between visitor home zip code and campsite
zip_code_population.x	acs	the zip code population when get_acs() from tidycensus pu
asian	acs	estimated percentage of asian population in a zip code
black	acs	estimated percentage of black population in a zip code
hispanic_latinx	acs	estimated percentage of hispanic latinx population in a zip
multiracial	acs	estimated percentage of multiracial population in a zip cod
native_american	acs	estimated percentage of native american population in a zi
other	acs	estimated percentage of other population in a zip code
pacific_islander	acs	estimated percentage of pacific islander population in a zip
white	acs	estimated percentage of white population in a zip code
zip_code_population.y	acs	the zip code population when get_acs() from tidycensus pu
college	acs	estimated percentage of population with a college degree in
hs_GED_or_below	acs	estimated percentage of population with a high school Gen
master_or_above	acs	estimated percentage of population with a master degree o
some_college	acs	estimated percentage of population with some college educa
median_income	acs	median household income in the past 12 months in 2019 in
zip_code_population	acs	total number of people living in a zip code when get_acs()
english_only	acs	estimated percentage of population that speak only english
not_english_only	acs	estimated percentage of population that speak a language o

Chapter 7

Summary of Testing

7.1 Data Integrity

We screened the data for outliers by summarizing and visualizing the raw data, and assessed whether those outliers need to be removed. We consulted other researchers who are familiar with the RIDB datasets to confirm outliers or other anomalies in the data.

Additionally, we documented the percent loss from data wrangling to ensure that our cleaning and wrangling of the data were reasonable.

7.2 Code Review

We conducted code reviews within the team, and with faculty or external advisers. We reviewed specific code chunks and scripts related to the Outdoor Equity App.

We separated our workflows so that one person created scripts, and the other reviewed them. We did this to maintain some objectivity when evaluating if our datasets were aggregating correctly. We also had a separate workflow for metadata, where one person created and wrote metadata, and the other reviewed it. This confirmed that the data matches how it is being described in the metadata. This confirmation is important as we want our client to be able to scale our product and workflows for future use.

7.3 Product Testing

We used three packages to test our R Shiny app. We used `shinytest` [Chang et al., 2021b] to ensure our app is visualized the way we expect it to using the package's snapshot-based testing strategy. We used `shinyloadtest` [Schloerke et al., 2021] to test the server hosting the R Shiny App to ensure that it responds in a reasonable amount of time based on the inputs a user provides. Similarly, we utilized the `tictoc` [**?**] package during our data wrangling and cleaning, and when we initially created our plots, graphs, and maps to estimate an informed guess of how long it may take the app to run our scripts. Lastly, we used the `reactlog` [Schloerke, 2020] package's diagnostic tool which creates a reactive visualizer for the app to make sure that reactive elements are working the way we expected them to. It is important to note that this diagnostic tool was not useful as our app functionality increased, as the reactive visualizer became impossible to read. There may be other options within `reactlog` to use the reactive visualizer in a different way, but we did not have enough time to research this.

We added temporary print statements to all functions in the app to ensure that the functions were working correctly were are outputting what we expect. We removed print statements from functions that were functioning with zero errors. We did this because print statements can take a long time to run and should not be left in functions or in the app permanently.

Additionally, we held multiple meetings with the Recreation One Stop team to obtain real-time and focused feedback to improve user design and experience.

7.3.1 Next steps for testing

Due to time constraints, we were not able to implement all testing methods we wanted. We recommend the following testing strategies to make the app more robust and for smoother functionality.

- Use the R package `testthat` [Wickham, 2022] to conduct unit tests on the scripts used to create Tidy datasets and for subsetted datasets for visualization. This type of testing may be important to avoid silent failures and to ensure that the datasets are aggregating correctly.
- Use `gremlin.js`), a JavaScript library used for "Monkey testing" to test the behavior of the R Shiny App. This package is compatible with `shiny` [Chang et al., 2021a] and does not require any external installation. See Chapter 11 in Engineering Production-Grade Shiny Apps for more guidance. "Monkey testing" is a type of testing where random, automated tests provide random inputs and then checks the behavior of the app (i.e. if the system or application crashes). We were able to find some finicky bugs through our own testing of random inputs, but "Monkey testing" is the formal process of testing app behavior.

- Employ user testing with federal public land managers, researchers, and those who are not familiar with RIDB data to further enhance the user experience and design.

Chapter 8

User Documentation

8.1 Purpose of the Outdoor Equity App

The Outdoor Equity App provides users with tools to explore trends at different overnight reservable sites to analyze access to these sites. The intended audience is federal public land managers and researchers as well as nonprofit organizations and recreation users.

The Recreation Information Database (RIDB) data are comprehensive when it comes to information regarding the site and reservation, but do not include information about the visitor outside of their home ZIP code. US Census American Community Survey (ACS) data is used to approximate socioeconomic demographics by joining information about the visitors' home ZIP code to the RIDB data.

8.2 How to Use the Outdoor Equity App

8.2.1 About the App

The About tab of the Outdoor Equity App includes background information about what the App is, why outdoor recreation is important, the App creators, and what data is used in the App. These sections are similar to the About Section of this technical documentation. The About tab also includes example questions that a user might explore through the different parts of the Analysis tab.

The Analysis tab of the App consists of three parts Data Summary, Data Relationship and Visitorshed Maps. Each of these pages includes a brief explanatory section of how to interpret the plot or graph, a section to subset the data to the desired campsite, and the plot or map outputs.

The Metadata tab of the App includes metadata for all variables in the combined RIDB and ACS dataset. This section mirrors the Metadata section in the Products and Deliverables Section of this document.

The Data Download section of the App allows a user to download a subset of data include as many or as few campsites and variables as they require for further analysis or use.

8.3 How to Maintain the Outdoor Equity App

8.3.1 Data Preparation Methods

8.3.1.1 RIDB Data

RIDB data data are available through direct download as CSV files or via the API as JSON files via Recreation.gov. API access requires creating a Recreation.gov account and requesting second tier API access via the Recreation.gov website's Contact Us page. CSV files are readily available for download via the Recreation.gov website. Data are collected each time a visitor makes a reservation through Recreation.gov. Data packages are posted annually in the spring by R1S and contain the previous fiscal year's reservations (ex: the 2018 package includes 2018-10-01 through 2018-09-30). Data packages are available for download from 2006 to present. Each annual data package file contains a range between 2 million and 5 million observations (or reservations) and includes variables in character, numeric, and date/time formats about each reservation. A shift in the data collection and storage processes occurred in 2019, changing what variables are available and how they are labeled. Currently, the Outdoor Equity App contains only data for reservable sites in California and reservations for fiscal year 2018.

A reproducible workflow for cleaning and wrangling data is employed in the `data_wrangle_and_clean.Rmd` document that sources custom functions to prepare RIDB data for joining with ACS data. All custom functions locations within the repository are listed in the Access, Clean, and Wrangle Data Section of this document. These functions rely heavily on functions from the `tidyverse` collection of packages [Wickham, 2021].

First, a the `function_ridb_subset-pre2018.R` is used to subset the RIDB data, filtering only reservations within the selected state that are listed as "Overnight" reservations within the `use_type` variable. For the 2018 California dataset this results in a starting "raw" data frame with 521,682 reservations

```r
# filter for state
filter(facility_state == state_abbrev) %>%
  # filter for use type
  filter(use_type == "Overnight")
```

The function then selects only the necessary variables, including information about the site (agency, park or forest name, site name, site type, and site location) and information about the reservation (home ZIP come, total paid, visit start and end dates, visit order date, and number of people in party).

```r
# select variables
select(c("agency", "parent_location", "region_description", "park",
         "site_type", "facility_id", "facility_state", "facility_longitude",
         "facility_latitude", "customer_zip", "total_paid", "start_date",
         "end_date", "order_date", "number_of_people")) %>%
  mutate(site_type = tolower(site_type)) %>%
  filter(!site_type %in% c("historic tour", "hiking zone", "group picnic area", "cave tour",
                           "management", "picnic", "entry point", "trailhead"))
```

The customer ZIP code values are then normalized. This includes filtering for only US ZIP codes and shortening all 9 digit ZIP codes to include only the first 5 digits.

```r
# filter out invalid ZIP codes
filter(str_detect(string = customer_zip,
                  pattern = "^[:digit:]{5}(?!.)") |
         str_detect(string = customer_zip,
                    pattern = "^[:digit:]{5}(?=-)")) %>%
  filter(!customer_zip %in% c("00000", "99999")) %>%
  mutate(customer_zip = str_extract(string = customer_zip,
                                    pattern = "[:digit:]{5}"))
```

This function results in the removal of 24,866 reservations (or 4.77%) from the original "raw" dataset that included all reservations for reservable overnight campsites in California.

A second custom function is then utilized to calculate and manipulate variables of interest. Start, end, and order dates calculate the lengths of stay and booking windows (number of days from order to start date) of each reservation. The booking window calculations return a number of results that are negative. This is a known issue that others working with the RIDB data have encountered. This resulted in 4,530 reservations (or 0.87%) without a valid booking window.

```
mutate(start_date = as.Date(start_date),
       end_date = as.Date(end_date),
       order_date = as.Date(order_date),
       # calculate new variables
       length_of_stay = as.numeric(difftime(end_date, start_date), units = "days"),
       booking_window = as.numeric(difftime(start_date, order_date), units = "days"))
```

Total costs are divided by lengths of stay to calculate cost per day and cost per day per visitor.

```
# calculate new variables
mutate(daily_cost = total_paid / length_of_stay,
       daily_cost_per_visitor = daily_cost / number_of_people)
```

Site types are aggregated to create 7 broader site categories.

```
# aggregate site type
mutate(aggregated_site_type =
         case_when(site_type %in% c("walk to", "hike to", "group hike to", "group walk
         ) ~ "remote",
         site_type %in% c("cabin nonelectric", "cabin electric", "yurt","shelter nonel
         ) ~ "shelter",
         site_type %in% c("boat in", "anchorage") ~ "water",
         site_type %in% c("group equestrian", "equestrian nonelectric") ~ "equestrian"
         site_type %in% c("rv nonelectric", "rv electric", "group rv area nonelectric"
         ) ~ "rv only",
         site_type %in% c("group standard nonelectric", "standard nonelectric",
                          "standard electric", "group standard area nonelectric",
                          "group standard electric"
         ) ~ "rv or tent",
         site_type %in% c("tent only nonelectric", "group tent only area nonelectric",
                          "tent only electric"
         ) ~ "tent only"))
```

The administrative unit variable is created by combining the `parent_location` and `region_description` variables as different federal agencies track the administrative unit information in different variables. Then the `agency`, `admin_uni` and `park` variables character strings are updated using multiple functions from the `stringr` package [Wickham, 2019].

```
mutate(admin_unit = case_when(agency == "USFS" ~ parent_location,
                              agency %in% c("NPS", "BOR", "USACE") ~ region_descriptio
  # edit values
  mutate(
```

```
# agency abbreviations to names
agency = str_replace(string = agency,
                     pattern = "NPS",
                     replacement = "National Park Service"),
agency = str_replace(string = agency,
                     pattern = "USFS",
                     replacement = "US Forest Service"),
agency = str_replace(string = agency,
                     pattern = "USACE",
                     replacement = "US Army Corps of Engineers"),
agency = str_replace(string = agency,
                     pattern = "BOR",
                     replacement = "Bureau of Reclamation"),
# update admin_unit values (generic)
admin_unit = str_replace(string = admin_unit,
                         pattern = paste(c("NF - FS", "NF -FS", "NF- FS",
                                           "NF-FS", "-FS", " - FS"),
                                         collapse = "|"),
                         replacement = "National Forest"),
admin_unit = str_to_title(admin_unit),
admin_unit = str_replace(string = admin_unit,
                         pattern = "And",
                         replacement = "&"),
# update park values (generic)
park = str_remove(string = park,
                  pattern = paste(c("\\(.*", " \\(.*",
                                    "---.*", " ---.*",
                                    ",.*"), collapse = "|")),
park = str_to_title(park),
park = str_replace(string = park,
                   pattern = "Cg",
                   replacement = "Campground"),
park = str_replace(string = park,
                   pattern = "Nhp",
                   replacement = "National Historic Park"),
park = str_replace(string = park,
                   pattern = "@",
                   replacement = "At"),
park = str_replace(string = park,
                   pattern = "&",
                   replacement = "And"),
park = str_replace(string = park,
                   pattern = paste(c("/", " / "), collapse = "|"),
                   replacement = " "),
park = str_remove_all(string = park,
```

```
                          pattern = " \\d.*"),
    # update park values (CA specific)
    park = str_remove(string = park,
                      pattern = paste(c(" - Angeles Nf", " -Hwy"),
                                      collapse = "|")),
    park = str_replace(string = park,
                       pattern = "Tunnel Mills Il",
                       replacement = "Tunnel Mills"),
    admin_unit = str_replace(string = admin_unit,
                             pattern = "Sequoia & Kings Canyon National Park$",
                             replacement = "Sequoia & Kings Canyon National Parks"))
```

Distance traveled is calculated by measuring the distance from the latitude and longitude facility coordinate locations to the centroid of the home ZIP code, which is accessed via the `tidycensus` package [Walker and Herman, 2022].

```
# bootstrap geometries and reproject to NAD 83
df_geometries <- df %>%
  st_as_sf(coords = c("facility_longitude", "facility_latitude"),
           crs = 4326) %>%
  st_transform(crs = 4269) # using NAD83 because measured in meters

# get centroid of geometries for all US ZIP codes
df_zip_centroids_us <- get_acs(geography = "zcta", year = 2018, geometry = TRUE,
                               summary_var = "B01001_001",
                               survey = "acs5",
                               variables = c(male = "B01001_002")) %>%
  select(NAME, geometry) %>%
  mutate(zip_code = str_sub(NAME, start = -5, end = -1)) %>%
  select(zip_code, geometry) %>%
  st_centroid()

# join data and calculate `distance_traveled` variable
df_joined_geometries <-
  left_join(x = df_geometries %>% as.data.frame(),
            y = df_zip_centroids_us %>% as.data.frame(),
            by = c("customer_zip" = "zip_code")) %>%
  st_sf(sf_column_name = 'geometry.x') %>%
  mutate(distance_traveled_m = st_distance(x = geometry.x,
                                           y = geometry.y,
                                           by_element = TRUE),
         distance_traveled_m = as.numeric(distance_traveled_m))
```

And finally, a variable is added indicating in which state or territory each customer zip code is located. This portion of the code utilizes the `zipcodeR` package

[Rozzi, 2022].

```
# create df of fips and full state names
fips_list <- c("01", "02", "04", "05", "06", "08", "09", "10", "11", "12",
               "13", "15", "16", "17", "18", "19", "20", "21", "22", "23",
               "24", "25", "26", "27", "28", "29", "30", "31", "32", "33",
               "34", "35", "36", "37", "38", "39", "40", "41", "42", "44",
               "45", "46", "47", "48", "49", "50", "51", "53", "54", "55",
               "56", "72")
state_list <- c("AL", "AK", "AZ", "AR", "CA", "CO", "CT", "DE", "DC", "FL",
                "GA", "HI", "ID", "IL", "IN", "IA", "KS", "KY", "LA", "ME",
                "MD", "MA", "MI", "MN", "MS", "MO", "MT", "NE", "NV", "NH",
                "NJ", "NM", "NY", "NC", "ND", "OH", "OK", "OR", "PA", "RI",
                "SC", "SD", "TN", "TX", "UT", "VT", "VA", "WA", "WV", "WI",
                "WY", "PR")
states_names_list <- c("Alabama", "Alaska", "Arizona", "Arkansas", "California",
                       "Colorado", "Connecticut", "Delaware", "District of Columbia",
                       "Florida","Georgia", "Hawaii", "Idaho", "Illinois", "Indiana",
                       "Iowa", "Kansas", "Kentucky", "Louisiana", "Maine","Maryland",
                       "Massachusetts", "Michigan", "Minnesota", "Mississippi", "Missouri",
                       "Montana", "Nebraska", "Nevada", "New Hampshire","New Jersey",
                       "New Mexico", "New York", "North Carolina", "North Dakota", "Ohio",
                       "Oklahoma", "Oregon", "Pennsylvania", "Rhode Island","South Carolina",
                       "South Dakota", "Tennessee", "Texas", "Utah", "Vermont", "Virginia",
                       "Washington", "West Virginia", "Wisconsin","Wyoming", "Puerto Rico")
df_states_fips <- as.data.frame(list(fips = fips_list,
                                     state = state_list,
                                     state_full = states_names_list))

# loop through state df to get all ZIP codes w/in state
df_states_zip_codes <- data.frame()

for (i in seq_along(fips_list)){
  state <- zipcodeR::search_fips(state_fips = fips_list[[i]]) %>%
    select(zipcode, state)
  df_states_zip_codes <- rbind(df_states_zip_codes, state)
}

# add full state name and fips code to list of all ZIP codes for each state
df_states_fips_zip_codes <-
  left_join(x = df_states_zip_codes,
            y = df_states_fips,
            by = "state") %>%
  select(-fips) %>%
  rename(customer_zip_state = state,
         customer_zip_state_full = state_full,
```

```
        zip_code = zipcode)
```

8.3.1.2 U.S. Census Data

US Census data is publicly accessible in many ways. Our product utilizes the R package `tidycensus` [Walker and Herman, 2022] to access the necessary variables from the 2018 American Community Survey (ACS) via API. API access requires an api key, which can be .

```
# API set up
# ONLY HAVE TO RUN THE FIRST TIME USING THIS RMD on a new machine
census_api <- source("private/census-api.R")
census_api_key(key = census_api[[1]], install = TRUE, overwrite = TRUE)
# run in console:
readRenviron("~/.Renviron")

# look at option variables
#View(load_variables(2018, "acs5", cache = TRUE))
```

Sample data are collected for the ACS each year and includes many variables that cover social, economic, housing, and demographic characteristics. The Outdoor Equity App utilizes the ACS 5-year data, which is an estimate representing data collected over the designated 5 year period. We used the 5-year ACS data over the 1-year ACS data because it increases "statistical reliability of the data for less populated areas and small population subgroups" [Bureau, 2022].

The variables included in the App include median-income, race, language(s) spoken at home, and highest level of education attained. All variables are represented as estimates in numeric format for a ZIP code tabulation area. Data are called by geographic region, in our case the ZIP code tabulation area, and include an estimated number of people that fall into each category within each ZIP code, a margin of error, and an estimated number of total people in the area. Within our custom functions `function_acs_education.R`, `function_acs_language.R`, `function_acs_median_income.R`, and `function_acs_race.R` we first imported just the necessary columns for each ACS variable.

```
# import variables for race
acs_df <-
  get_acs(geography = "zcta", year = 2018,
          state = "CA", survey = "acs5",
          summary_var = "B03002_001", #Estimate!!Total:
          variables = c(
            white = "B03002_003", #Estimate!!Total!!Not Hispanic or Latino!!White alon
```

```
              black = "B03002_004", # Estimate!!Total:!!Not Hispanic or Latino!!Black or African Am
              native_american = "B03002_005", # Estimate!!Total:!!Not Hispanic or Latino!!American
              asian = "B03002_006", # Estimate!!Total:!!Not Hispanic or Latino!!Asian alone
              pacific_islander = "B03002_007", # Estimate!!Total:!!Not Hispanic or Latino!!Native H
              other = "B03002_008", # Estimate!!Total:!!Not Hispanic or Latino!!Some other race alo
              multiracial = "B03002_009", # Estimate!!Total:!!Not Hispanic or Latino!!Two or more r
              hispanic_latinx = "B03002_012" # Estimate!!Total!!Hispanic or Latino
            ))
```

```
## Getting data from the 2014-2018 5-year ACS
```

```
head(acs_df, 10)
```

```
## # A tibble: 10 x 7
##    GEOID NAME        variable         estimate   moe summary_est summary_moe
##    <chr> <chr>       <chr>               <dbl> <dbl>       <dbl>       <dbl>
##  1 90001 ZCTA5 90001 white                 413   165       58975        1725
##  2 90001 ZCTA5 90001 black                5138   646       58975        1725
##  3 90001 ZCTA5 90001 native_american        37    42       58975        1725
##  4 90001 ZCTA5 90001 asian                  67    38       58975        1725
##  5 90001 ZCTA5 90001 pacific_islander        0    29       58975        1725
##  6 90001 ZCTA5 90001 other                 168   169       58975        1725
##  7 90001 ZCTA5 90001 multiracial            66    44       58975        1725
##  8 90001 ZCTA5 90001 hispanic_latinx     53086  1591       58975        1725
##  9 90002 ZCTA5 90002 white                 223    81       53111        2031
## 10 90002 ZCTA5 90002 black               10110   876       53111        2031
```

Percentages for the categories within each ACS variable are calculated for race, language(s) spoken in the home, and highest level of education by dividing the estimate for a category by the estimated total population for that ZIP code.

For median-income, the estimated median-income is used as is, without further adjustments.

ACS data are then pivoted wider, moving the different categories and percentages (ex: racial groups) to their own columns to create a single row for each ZIP code. This is necessary for joining the ACS data sets to the RIDS data set.

8.3.1.3 Data Joining

8.3.2 Statistical Analysis and Data Wrangling for Plots

Each type of plot and map require unique data wrangling which is explained in this section.

8.3.2.1 Data Summary

8.3.2.2 Data Relationships

8.3.2.3 Spatial analysis

8.3.3 Data Limitations

8.3.4 Data updates

8.3.5 Server Hosting

8.3.6 Shiny Code Directory

8.4 How to Expand the Outdoor Equity App

8.4.1 Temporal Expansions

8.4.2 Spatial Expansions

8.4.3 Statistical Analysis

Chapter 9

Additional Challenges

9.1 Technical Challenges

- In the Outdoor Equity App, there are behavioral bugs with the plots where plots will stop updating or being created after either an input has been changed multiple times or if the input chosen is associated with a large amount of data. We were not able to troubleshoot this completely, but believe this has to do with the functions that create the visuals.

9.2 Future Challenges

- Adding in multiple years and adding in more spatial data (beyond California) will require a large server and likely more computing power.

Chapter 10

Archive Access

Bibliography

John C. Bergstrom, Matthew Stowers, and Scott J. Shonkwiler. What does the future hold for u.s. national park visitation? estimation and assessment of demand determinants and new projections. 45(1):38 – 55, 2020. doi: 10.22004/AG.ECON.298433.

United States Census Bureau. American community survey 5-year data (2009-2020). website, 2022. URL: https://www.census.gov/data/developers/data-sets/acs-5year.html.

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. *shiny: Web Application Framework for R*, 2021a. URL https://shiny.rstudio.com/. R package version 1.7.1.

Winston Chang, Gábor Csárdi, and Hadley Wickham. *shinytest: Test Shiny Apps*, 2021b. URL https://github.com/rstudio/shinytest. R package version 1.5.1.

Alan Ewert and Steve Hollenhorst. Resource allocation: Inequities in wildland recreation. 61(8):32 – 36, 1990. doi: 10.1080/07303084.1990.10604598.

David Flores, Gennaro Falco, Nina S. Roberts, and III Francisco P. Valenzuela. Recreation equity: Is the forest service serving its diverse publics? 116(3): 266 – 272, 2018. doi: 10.1093/jofore/fvx016.

Myron Floyd and Cassandra Y. Johnson. Coming to terms with environmental justice in outdoor recreation: A conceptual discussion with research implications. 24(1):59 – 77, 2002. doi: 10.1080/01490400252772836.

William E. Hammitt, David N. Cole, and Christopher A. Monz. *Wildland Recreation: Ecology and Management.* John Wiley and Sons, Oxford, UK, 2015.

E.P. Meinecke. Recreation planning: A discussion. 35(12):1120 – 1128, 1937. doi: 10.1093/jof/35.12.1120.

William L. Rice and So Young Park. Big data spatial analysis of campers' landscape preferences: Examining demand for amenities. 292(15), 2021. doi: 10.1016/j.jenvman.2021.112773.

William L. Rice, So Young Park, Bing Pan, and Peter Newman. Forecasting campground demand in us national parks. 75:424 – 438, 2019. doi: 10.1016/j.annals.2019.01.013.

Gavin Rozzi. *zipcodeR: Data & Functions for Working with US ZIP Codes*, 2022. https://github.com/gavinrozzi/zipcodeR/, https://www.gavinrozzi.com/project/zipcoder/.

Joseph L. Sax. *Mountains Without Handrails: Reflections on the National Parks*. University of Michigan Press, Ann Arbor, MI, 1980.

Barret Schloerke. *reactlog: Reactivity Visualizer for shiny*, 2020. URL https://CRAN.R-project.org/package=reactlog. R package version 1.1.0.

Barret Schloerke, Alan Dipert, and Barbara Borges. *shinyloadtest: Load Test Shiny Applications*, 2021. URL https://CRAN.R-project.org/package=shinyloadtest. R package version 1.1.0.

David Scott and KangJae Jerry Lee. People of color and their constraints to national parks visitation. 35(1):73 – 82, 2018.

Mostafa Shartaj and Jordan F. Suter. Exploring the local determinants of campground utilization on national forest land. 18(2):114 – 128, 2020. doi: 10.22004/ag.econ.308121.

Bo Shelby, Doug Whittaker, and Mark Danley. Idealism versus pragmatism in user evaluations of allocation systems. 11(1):61 – 70, 1989. doi: 10.1080/01490408909512205.

Abby L. Timmons. Too much of a good thing: Overcrowding at america's national parks. 94(2):985 – 1017, 2019.

Kyle Walker and Matt Herman. *tidycensus: Load US Census Boundary and Attribute Data as tidyverse and sf-Ready Data Frames*, 2022. URL https://walker-data.com/tidycensus/. R package version 1.2.1.

Margaret Walls, Casey Wichman, and Kevin Ankney. Nature-based recreation: Understanding campsite reservations in national parks. Technical report, Resources for the Future, 2018.

Hadley Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*, 2019. URL https://CRAN.R-project.org/package=stringr. R package version 1.4.0.

Hadley Wickham. *tidyverse: Easily Install and Load the Tidyverse*, 2021. URL https://CRAN.R-project.org/package=tidyverse. R package version 1.3.1.

Hadley Wickham. *testthat: Unit Testing for R*, 2022. URL https://CRAN.R-project.org/package=testthat. R package version 3.1.2.

Xiao Xiao, KangJae Jerry Lee, and Lincoln R. Larson. Who visits u.s. national parks (and who doesn't)? a national study of perceived constraints and vacation preferences across diverse populations. 53(3):404 – 425, 2021. doi: 10.1080/00222216.2021.1899776.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2021. URL https://CRAN.R-project.org/package=bookdown. R package version 0.24.