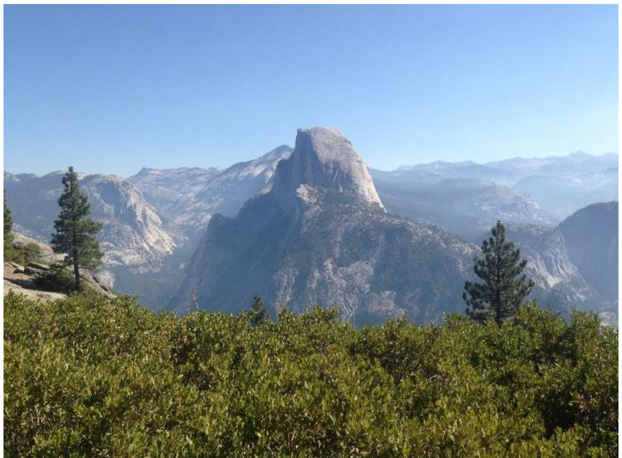

National Parks Recreational Database Project

LOCATION

Assateague Island National Park
Grand Teton National Park
Great Smoky Mountain National Park
Rocky Mountain National Park

TYPE OF WORK

Recreational Database Creation

Final Project Report – Group 1

Prepared by:

Michael Connolly (9083777145),
Kuang-Cheng Cheng (9083544867),
Drew Smith (9077663509)

Prepared for:

Geography 574,
Professor Qunying Huang

University of Wisconsin - Madison
Science Hall
Department of Geography

December 19, 2022

This page intentionally left blank.

Contents

Abstract ii

1 Introduction 1

1.1 Problem Statement / Objective 1

1.2 Datasets 1

2 Database Description 3

2.1 Conceptual Design 3

2.2 Logical Design 3

3 Methodology and Implementation 6

3.1 Non-Spatial Data 6

3.2 Spatial Data 8

3.3 Discussion of Implementation 10

4 Results 12

4.1 Scenarios, Questions, and Queries 12

4.2 Conclusions and Future Efforts 15

5 References Cited 16

Tables

Table 1-1. Research Questions 1

Table 1-2. Data Sources and Links 2

Figures

Figure 2-1. National Park Database Entity Relationship Diagram created in ERDPlus. 4

Figure 2-2. National Park Database Relational Schema created in ERDPlus. 5

Figure 3-1. Data excerpts depicting the online view on the left and downloaded CSV format on the right (NPS 2022b). 6

Figure 3-2. Coding to alter the table fields and add national park names. 7

Figure 3-3. Coding to merge the separate national park documents into a single CSV file. 7

Figure 3-4. Coding to identify data length. 8

Figure 4-1. Campgrounds, historical sites, trails, and parking areas within Rocky Mountain National Park. 14

Abstract

Currently there are 423 national parks in the United States containing various amenities such as hiking trails, campgrounds, and historical places of interest. According to recent data, overall national park visitation is approximately 297 million visitors per year with only 25 national parks accounting for 50 percent of overall visitation (NPS 2022a). Severe overcrowding in the most popular national parks may detract or ruin the experience for many people. Additionally, parking and access may be a potential consideration for some visitors. Therefore, the objective of this project is to provide visitors search tools so they can determine which national parks they would like to visit based on their personal interests and possible concern of overcrowded parks or parks with limited availability and access.

This database was designed for use with pgAdmin and five questions were devised to demonstrate its functionality and versatility. Formulation of these questions was based on a hypothetical scenario of a user planning a visit to a national park and searching for various aspects that could affect their decision to visit one park over another. The datasets used to construct this database were obtained from the National Park Service (NPS) and the United States Department of Agriculture, Forest Service (USDA-FS). Targeted features included park visitation data, roads, trials, parking areas, historical points of interest, and facilities such as campgrounds.

This report details the process of designing a national parks database from conceptualization to implementation. As the data obtained for the database was fragmented or incomplete, several issues were encountered that were dealt with throughout the process such as each national park using different spatial data structures. Using ArcGIS Pro, Microsoft Excel, Jupyter Notebook, and pgAdmin, the data was altered, edited, and combined to create a more cohesive dataset that can be queried using PostgreSQL. Although this database is an improvement over existing databases, the addition of spatial data for more national parks is a target for future efforts.

1 Introduction

1.1 Problem Statement / Objective

Currently there are 423 national parks in the United States with various amenities such as hiking trails, campgrounds, and historical places of interest. Approximately 25 of these national parks account for 50 percent of overall visitation (NPS 2022a). Severe overcrowding and access issues in some national parks may detract or ruin the experience for many people. Therefore, the objective of this project is to provide visitors search tools so they can determine which national parks they would like to visit based on their personal interests and concerns to potentially avoid overcrowded parks or parks with limited availability and access.

Example research questions that can be answered with the proposed database are included in **Table 1-1** below and discussed further in Section 4. These questions represent a variety of possible queries from the national parks database.

Table 1-1. Research Questions

Research Questions	
Q1.	What National Parks are in Colorado and what are the sizes of each park?
Q2.	What was the total visitation of each national park from 2010 to 2020?
Q3.	What are the total number of campsites, historical sites, trails, and parking areas in Rocky Mountain National Park?
Q4.	How many unpaved roads are in Rocky Mountain, Great Smoky Mountain, AND Grand Teton National Parks, and what are the total lengths of the roads?
Q5.	Which National Park has visitation larger than the average visitation of all four National Parks in this database?

1.2 Datasets

The datasets used to construct this database were obtained from the National Park Service (NPS) and the United States Department of Agriculture, Forest Service (USDA-FS) (**Table 1-2**). Visitation data, park boundaries, roads, trails, parking areas, and historical / cultural places of interest were obtained from the NPS. Visitation data was in comma separated values (CSV) format, and all other data obtained from the NPS were in shapefile format. The majority of the NPS data was separated by national park and had to be merged to accommodate the database design. Additionally, data for every feature class was not available for all national parks, so only select parks were included in the database tables for roads, trails, and parking areas. These tables only include Assateague Island, Grand Teton, Great Smoky Mountain, and Rocky Mountain National Parks.

Table 1-2. Data Sources and Links

Agency / Source	Description and Link
NPS (2022b)	Shapefiles for Park Boundaries, Roads, Trails, and Parking Areas. https://public-nps.opendata.arcgis.com/
NPS (2022c)	Shapefiles for the National Register of Historic Places (NRHP). https://www.nps.gov/subjects/nationalregister/database-research.htm
NPS (2022c)	CSV files for Visitation Data. https://irma.nps.gov/STATS/
USDA-FS (2022)	Shapefile for facilities ‐ campgrounds, campsites, trailheads, picnic areas, etc. https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=camp

* This information is also included in the references section.

2 Database Description

The conceptual and logical designs for the National Park Recreational Database were created using ERDPlus.

2.1 Conceptual Design

An entity relationship diagram was developed to illustrate the conceptualization of the database (**Figure 2-1**). The entities include National Parks represented by “nps_boundary”, Facilities such as campgrounds and other recreational amenities represented by “nps_facilities”, Historical Sites or Points of Interest represented by “nrhp_points”, Trails represented by “park_trails”, Roads represented by “park_roads”, Parking represented by “parking_areas”, and Visitors represented by “np_visitation”. Most of these entity relationships are one to many as one national park can have various quantities of the other entities. Additionally, all relationships are mandatory between the national parks and the other entities except the historical sites, which are not necessary components of a national park. The design was altered several times throughout the process to find a balance between what the database should contain and what data was available to realize the design.

2.2 Logical Design

A relational diagram (**Figure 2-2**) was created using the conceptual design / diagram as a basis. The various entities / tables were linked via primary and foreign keys – either “gid”, national park names, and/or “geometries”. There was difficulty in establishing some of these relationships due to duplicate or redundant fields in several datasets, such as the central “nps_boundary” dataset. Also, various software applications like ArcGIS Pro and pgAdmin automatically assigned unique identification numbers like Object_ID and “gid” respectively. Additionally, there were instances of different national park systems using their own identification numbers for entities and attributes. Therefore, it took several edit sessions to find the proper relationships and clean up the data. Prior to implementation, it was decided that the “gid” and “geometries” would be the primary links for the spatial data and park names for non-spatial data.

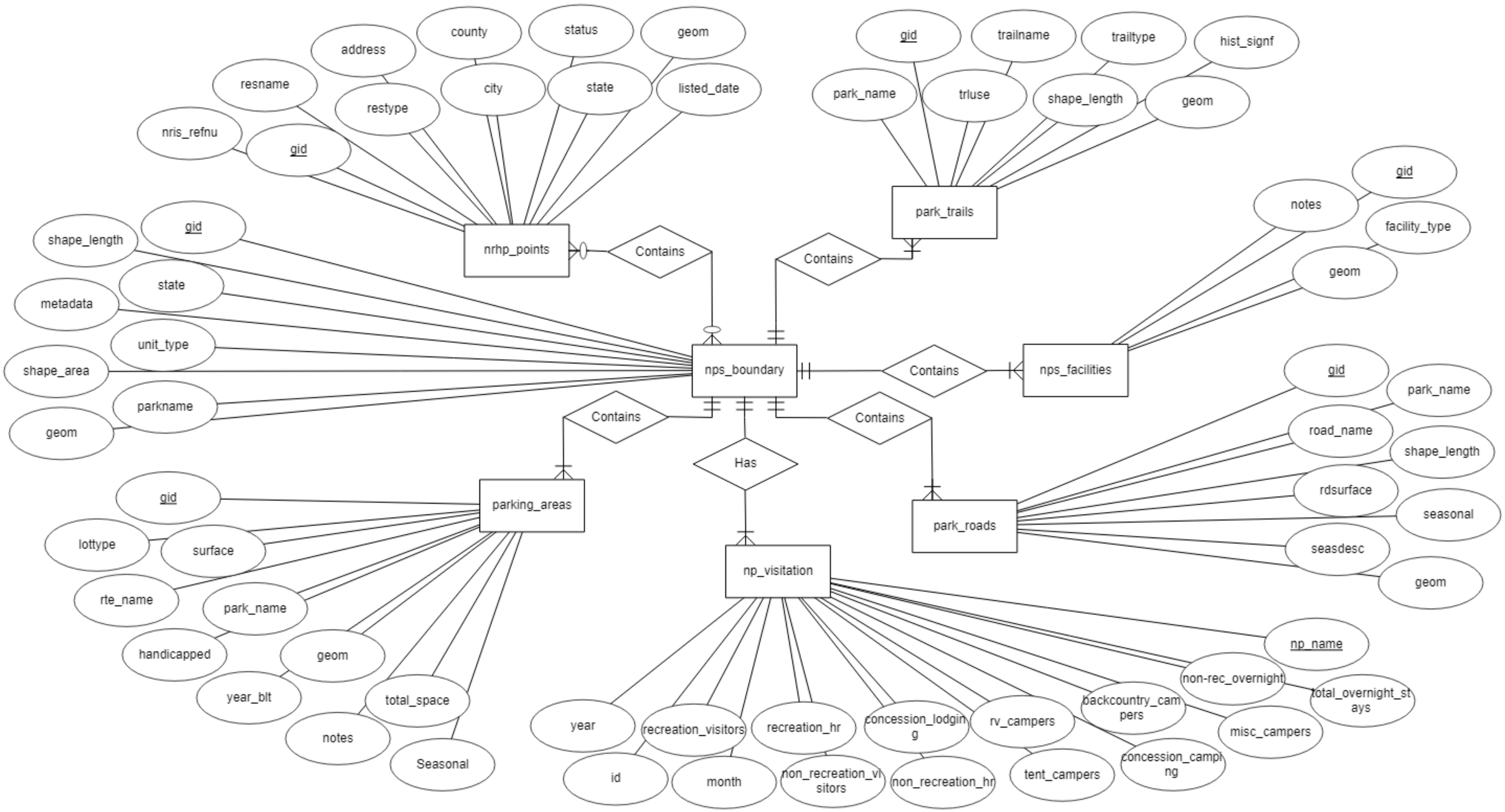

Figure 2-1. National Park Database Entity Relationship Diagram created in ERDPlus.

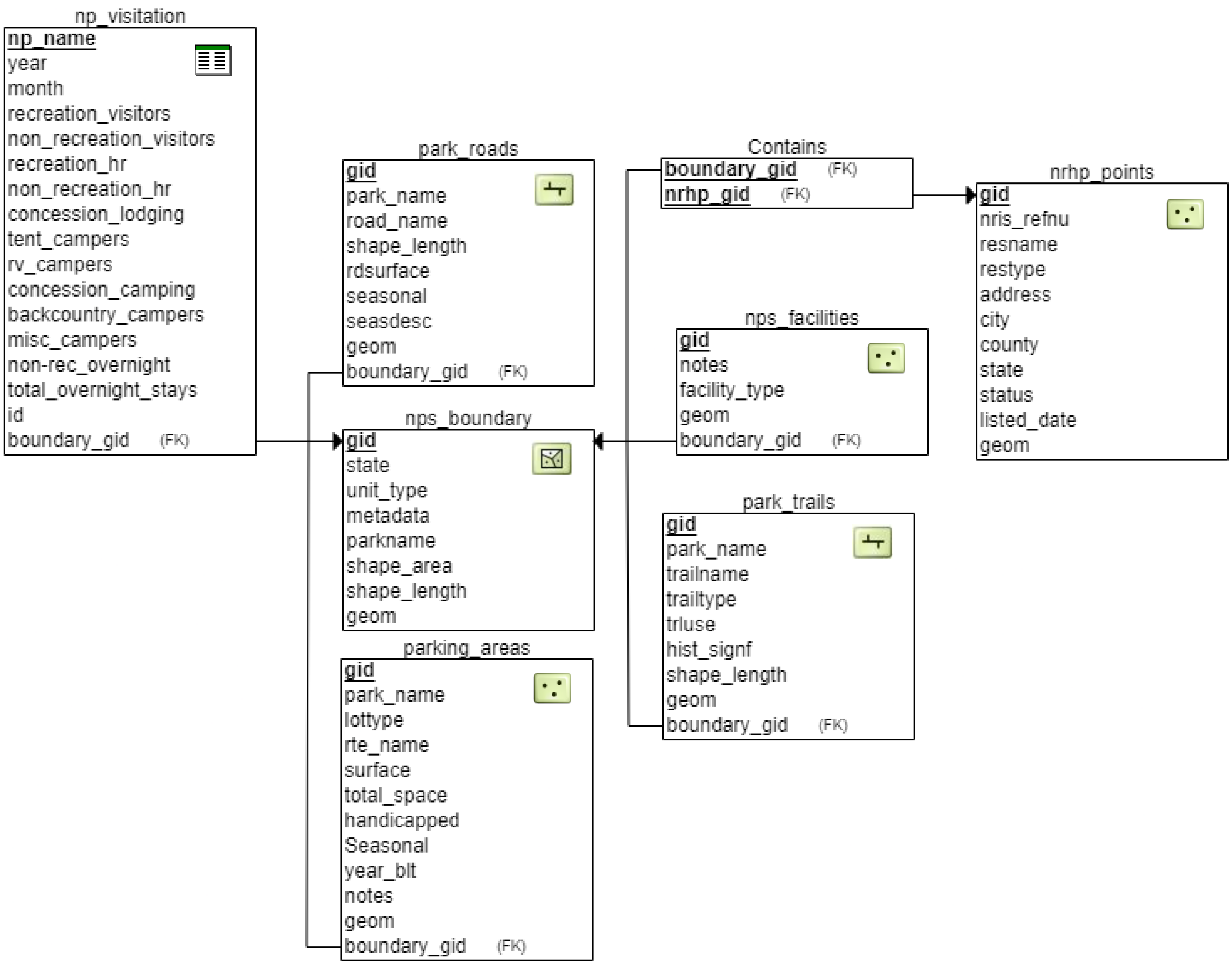

Figure 2-2. National Park Database Relational Schema created in ERDPlus.

3 Methodology and Implementation

This section discusses the methods used to prepare the data and enter it into pgAdmin (PostgreSQL). It is separated into three subsections: Non-Spatial Data, Spatial Data, and finally a discussion on coalescing the data into pgAdmin for query.

3.1 Non-Spatial Data

Four national park visitation datasets were used from the “Summary of Visitor Use By Month and Year (1979 - Last Calendar Year)” on the Integrated Resource Management Applications (IRMA) portal including Rocky Mountain National Park, Great Smoky Mountain National Park, Grand Teton National Park, and Assateague Island National Seashore. The datasets were in CSV file format and provided the total number of recreational visitors and non-recreational visitors by year and month, as well as the visitation for each type.

Non-Spatial Data Preparation

After the data was downloaded from the IRMA portal, the CSV file representation was different than what can be viewed online (**Figure 3-1**). All the field (column) names that were present in the online version were displayed as generic field names (i.e. Field 1, Field 2, etc.). Additionally, there were fields that were not applicable for the database, and therefore, excerpts were taken from the dataset and the field names were revised before importing the data into PostgreSQL.

	Recreation Visitors	Non-Recreation Visitors	Concession Lodging	Tent Campers	RV Campers	Concession Camping	Backcountry Campers
				2021			
January	140,715	90	0	198	186	0	63
February	133,771	50	0	186	273	0	44
March	155,819	80	0	435	351	0	65
April	183,792	70	0	1,278	1,311	0	141
May	335,060	490	0	3,915	2,562	0	1,974
June	671,232	1,400	0	11,589	6,144	0	6,835
July	904,767	2,050	0	12,606	5,985	0	11,939
August	746,343	2,300	0	12,447	5,784	0	10,226
September	521,233	1,500	0	8,730	5,445	0	6,184
October	352,736	560	0	3,627	2,805	0	1,760
November	164,992	100	0	708	948	0	103

Title1						
Assateague Island NS						
Field1	Field2	Field3	Field4	Recreatio	NonRecre	Field5
1979	January	48,050	200	288,360	100	0
1979	February	48,818	200	292,950	100	0
1979	March	50,794	200	324,874	100	0
1979	April	110,513	300	697,380	150	0
1979	May	150,960	350	1,240,904	250	0
1979	June	168,974	350	1,282,782	350	0
1979	July	271,543	350	2,351,722	350	0
1979	August	378,321	350	3,215,840	350	0
1979	Septembe	228,328	200	1,427,898	150	0
1979	October	102,143	200	620,982	150	0

Figure 3-1. Data excerpts depicting the online view on the left and downloaded CSV format on the right (NPS 2022b).

Visitation Data Processing (Using Pandas in Jupyter Notebook)

First, for processing convenience, all the CSV files were renamed using the official abbreviations with the years – "GRSM(1979 – 2021).csv" representing the data from Great Smoky Mountain for example. Second, the field names were revised to correspond with the correct field-name convention – "Field1":"Year", "Field2":"Month", etc., and a new field (column) name called "NP_name" was added as a label for each row of data. The new field was designed to hold the corresponding national park label for each row (**Figure 3-2**). Finally, the separate CSV files for each park were merged into a single table with 2,064 records (**Figure 3-3**).

```python
# Read csv file
import pandas as pd

# Enter the names
park_name_lst = ['Rocky Mountain', 'Great Smoky Mountains', 'Grand Teton', 'Assateague Island']
park_abbrev_lst = ['ROMO', 'GRSM', 'GRTE', 'ASIS']
lst_num = len(park_name_lst)
for i in range(lst_num):
    park_name = park_name_lst[i]
    park_abbrev = park_abbrev_lst[i]
    file_name = park_abbrev + '(1979 - 2021).csv'
    df = pd.read_csv(file_name, skiprows=3, usecols=range(14))

    # Revise Field Name
    df = df.rename(columns={"Field1":"Year", "Field2":"Month", "Field3":"Recreation Visitors", "Field4":"Non-Recreation Visitors",
    "Field5":"Concession Lodging", "Field6":"Tent Campers", "Field7":"RV Campers", "Field8":"Concession Camping",
    "Field9":"Backcountry Campers", "Field10":"Misc Campers"})

    df['NP_name'] = park_name

    # Export New CSV file
    export_name = park_abbrev + "_Visitation.csv"
    df.to_csv(export_name,index=False)
```

Figure 3-2. Coding to alter the table fields and add national park names.

```python
# Merge into one CSV file
export_name = "4NP_Visitatioin"
new_df = pd.concat(
    map(pd.read_csv, ['ROMO_Visitation.csv', 'GRSM_Visitation.csv', 'GRTE_Visitation.csv', 'ASIS_Visitation.csv']), ignore_index= True)

# Export Merged CSV
merged_name = '4NP_Visitation.csv'
new_df.to_csv(merged_name,index=False)
```

Figure 3-3. Coding to merge the separate national park documents into a single CSV file.

Preparation for Importing to PostgreSQL

To prepare the visitation data for import into PostgreSQL a new table had to be created to receive the data; however, defining the variables was the first step of importing. In order to minimize the usage of storage, we had to know the maximum length a field needs. There were two steps to find out the lengths and prepare the data to be imported:

(1) Make sure numbers were in numerical data type, and

(2) Search for the maximum length of each field (or column).

The first step was to change all the numbers in the CSV file into numerical data types; otherwise, they would be strings. If they were strings, they could not be imported into the table as numerical data types and calculations could not be performed during query. The second step was to find out the maximum length of each field (**Figure 3-4**). Although this function is not very accurate about the maximum lengths, it did give a rough estimate of the length for each field.

```
new_df = pd.read_csv('4NP_Visitation.csv')
new_df.max()
```

```
Year                                  2021
Month                            September
Recreation Visitors                1761918
Non-Recreation Visitors            1646888
RecreationHours                   13357493
NonRecreationHours                  823444
Concession Lodging                   69719
Tent Campers                         57561
RV Campers                          110645
Concession Camping                  122879
Backcountry Campers                  23216
Misc Campers                         17654
NonRecreationOvernightStays              0
TotalOvernightStays                 229646
NP_name                     Rocky Mountain
dtype: object
```

Figure 3-4. Coding to identify data length.

3.2 Spatial Data

The spatial data downloaded from NPS and USDA-FS was formatted differently depending on the specific source. Therefore, preparation of the data required fields to be parsed or merged, individual shapefiles to be merged and reprojected, and in some cases, the creation of new fields. Nearly every shapefile was in a different projection, so a new universal projection was chosen and applied to all shapefiles – "USA Contiguous Albers Equal Area Conic", GEOGCS ["NAD83", DATUM ["North American Datum 1983", SPHEROID "GRS1980".

National Park Facilities – Point Shapefile

The National Park Facilities Shapefile containing point locations of campgrounds, campsites, trailheads, and other facilities was downloaded as a complete file covering every national park and forest in the United States. However, the data necessary for the query was numerically coded and not useful for a text query. For example, campgrounds were represented as "63" and trailheads as "921". To prepare the data, the attribute table was exported to Excel from ArcGIS Pro and the codes were changed to the proper text name for all 21,904 records. The table was then converted to a CSV file and joined with the attributed table in ArcGIS Pro. Extraneous fields were deleted including any duplicate fields and those with no relevance such as "revision date".

NRHP Sites – Point Shapefile

The NRHP Sites shapefile containing historic points of interest was downloaded as a complete file with 67,443 records representing sites throughout the United States. The data was already formatted properly; however, there were many irrelevant fields or completely vacant fields. The data was examined carefully, and only fields pertinent to the current project were retained including official numerical designations, locational information, description, and status.

NPS Boundary – Polygon Shapefile

The NPS Boundary shapefile was downloaded with 428 records representing nearly every park and forest reserve in the United States. Like the NRHP Sites shapefile, the data in the attribute table was parsed so only relevant data was retained, such as "PARKNAME" and "Shape_Area". Additionally, forest reserves and duplicate entries were removed leaving 72 unique national parks in the dataset.

Roads and Trails – Line Shapefiles

The linear roads and trails shapefiles were only available for Assateague Island, Grand Teton, Great Smoky Mountain, and Rocky Mountain National Parks. Each national park had their own separate files for roads and trails, and the format of the tables was different for each park. The first step in preparation of these files was the removal of extraneous fields, then fields for different park datasets were compared and renamed if necessary to match the field names to the data. In some instances, data was transferred from one field to another for uniformity. Some of the tables were missing the field for "park_name", so a CSV file was created with the respective park names and then joined to the attribute tables. Finally, all the road tables and trail tables for each park were merged into cohesive datasets – one for roads with 4,315 records and one for trails with 1,152 records.

Parking Areas – Point Shapefile

The Parking Areas shapefile was in the same condition as the roads and trails shapefiles when downloaded from the NPS website. Each national park had its own data format with different field types. The data was parsed, reformatted, and combined to create a single Parking Areas Table with 872 records.

3.3 Discussion of Implementation

The following discussion has been separated by Non-Spatial and Spatial Data types as each datatype required different methods for import into pgAdmin. Additionally, this section discusses issues encountered and how they were resolved.

Non-Spatial Data

The following SQL statements were devised to create an empty table in pgAdmin for the national park visitation data and populate that table with data from the CSV file:

(1) Create empty table:

```
CREATE TABLE np_visitation(
    id SERIAL,
    year SERIAL,
    month VARCHAR(9),
    recreation_visitors     SERIAL,
    non_recreation_visitors SERIAL,
    recreation_hr SERIAL,
    non_recreation_hr SERIAL,
    concession_lodging SERIAL,
    tent_campers SERIAL,
    rv_campers SERIAL,
    concession_camping SERIAL,
    backcountry_campers SERIAL,
    misc_campers SERIAL,
    non_recreation_overnight_stays VARCHAR(1),
    total_overnight_stays SERIAL,
    np_name VARCHAR(21),
    PRIMARY KEY (id)
);
```

(2) Copy or import data from CSV to the empty table:

```
COPY
np_visitation(year,month,recreation_visitors,non_recreation_visitors,recreation_hr,
non_recreation_hr,concession_lodging,tent_campers,rv_campers,concession_camping,
backcountry_campers,misc_campers,
non_recreation_overnight_stays,total_overnight_stays,np_name)
FROM '< file path >\4NP_Visitation.csv'  --This is the path to the CSV file
DELIMITER ','CSV HEADER;
```

Two issues were encountered during the data import. The first issue involved the COPY query statement failing if the CSV file was stored on a cloud drive. There are two solutions for this issue. The first solution is to move the file to somewhere else in the computer, which is the local drive, to avoid using a cloud drive. The second solution is to use the PSQL Tool (or the command line) to copy the data using the following command:

```
\COPY
np_visitation(year,month,recreation_visitors,non_recreation_visitors,recreation_hr,no
n_recreation_hr,concession_lodging,tent_campers,rv_campers,concession_camping
,backcountry_campers,misc_campers,non_recreation_overnight_stays,total_overnig
ht_stays,np_name)
FROM '< file path >\4NP_Visitation.csv' DELIMITER ',' CSV HEADER
```

The reason for using the command line is that there are two functions that can copy data: copy and \copy; however, there is not a corresponding function (or syntax) for \copy in the query tool of pgAdmin4 so data on a cloud drive can only be accessed this way. The second issue was that querying data from the visitation table between two different years was not working. The first version of our empty table that was used to save the visitation data was unable to use “BETWEEN…AND…” command to search because the year was defined as “CHAR(4)” to save memory usage. However, it was discovered that “BETWEEN…AND…” can only work on numerical data. This was a simple fix by changing the data type for the “year” field from “CHAR(4)” to “SERIAL”.

Spatial Data

A spatial database was created in pgAdmin using UTF8 encoding and the postgis_32_sample as a template. The spatial data was then imported to pgAdmin using the “shp2pgsql-gui” PostGIS Shapefile Import/Export Manager. Initially, the data was projected in NAD 1983 HARN Contiguous USA Albers; however, when searching for a compatible spatial reference identifier (SRID) number, the choices provided in pgAdmin did not appear to be acceptable. Therefore, a new projection was chosen that did have an exact SRID match – 102003, USA Contiguous Albers Equal Area Conic.

After successfully reformatting and importing the shapefiles into the database, the data was examined for accuracy. It was discovered that many of the modifications made to the files during initial processing had reappeared. The issue was resolved by reexporting the modified shapefiles from ArcGIS Pro and importing the files into pgAdmin again. The data was reexamined after the second import and simple test queries were executed on each table to assess functionality. Further modifications were made to some tables using “ALTER TABLE”, “DROP COLUMN”, and “RENAME COLUMN” to either remove redundant columns, change impractical field names, or fix anomalies in field names that occurred during import.

4 Results

This section is separated into “Scenarios, Questions, and Queries” followed by “Conclusions and Future Efforts”. Scenarios, Questions, and Queries provides answers to the research questions introduced in Section 1 using sample queries. Conclusions and Future Efforts discusses the limitations of the database and potential improvements that can be made in the future.

4.1 Scenarios, Questions, and Queries

The following questions were formulated based on planning a hypothetical visit to a national park and each question and sample query is followed by a short discussion.

Question 1: What National Parks are in Colorado and what are the sizes of each park?

```
SELECT state, parkname, SUM(st_area(geom)) / 4047 park_acres
FROM nps_boundary
WHERE state = 'CO'
Group By parkname, state;
```

This question was designed from the point of view of someone looking for a national park to visit in Colorado. The query can be altered to include other states or specific national parks for comparison. Potential scenarios could be a user planning a trip to this region of the United States or possibly passing through the region on a road trip from Madison, WI to San Diego, CA.

Question 2: What was the total visitation of each national park from 2010 to 2020?

```
SELECT np_name, SUM(recreation_visitors + non_recreation_visitors) as visitors
FROM np_visitation
WHERE year BETWEEN 2010 AND 2020
GROUP BY np_name;
```

This question was designed from the point of view of a user wanting to examine visitation data of national parks to find which parks are consistently the most crowded over a given time. Some people would rather avoid overcrowded parks because too many people can distract from the experience. Visitation data was only compiled for four national parks, but of the four, it is clear that Great Smoky Mountains has approximately five times more visitors than the other parks. This query can be altered to aggregate monthly vitiation data and types of recreational visits. For instance, a user can find the number of back country campers in various parks during the summer months.

Question 3: What are the total number of campsites, historical sites, trails, and parking areas in Rocky Mountain National Park?

```
SELECT nb.parkname, COUNT(*) AS campground_num, nrhp.nrhp_sites AS nrhp_num,
trail_number.trails AS trail_num, parkinglot.parking AS pk_lot_num
FROM nps_boundary AS nb JOIN nps_facilities AS nf ON ST_Contains(nb.geom, nf.geom),
        (SELECT COUNT(*) AS nrhp_sites
        FROM nps_boundary AS b JOIN nps_facilities AS f ON ST_Contains(b.geom, f.geom)
        JOIN nrhp_points AS n ON ST_Contains(b.geom, n.geom)
        WHERE b.parkname = 'Rocky Mountain' AND f.facility_type = 'Campground'
        ) AS nrhp,
        (SELECT COUNT(*) AS trails
        FROM park_trails
        WHERE park_name = 'Rocky Mountain'
        ) AS trail_number,
        (SELECT COUNT(*) AS parking
        FROM parking_areas
        WHERE park_name = 'Rocky Mountain'
        ) AS parkinglot
WHERE nb.parkname = 'Rocky Mountain' AND nf.facility_type = 'Campground'
GROUP BY nb.parkname, nrhp_num, trail_num, pk_lot_num;
```

This question and query were designed to demonstrate the ability of the database to return several amenity types within a national park for comparison. The query will return the number of campgrounds, historical sites, trails, and parking areas in a national park (**Figure 4-1**). Queries can be modified to return the length of trails, the number of regular parking spaces and handicap spaces, and the number of scenic overlooks among other features. Furthermore, the historical sites can be searched for a particular site type and locate parks with that site type.

Question 4: How many unpaved roads are in Rocky Mountain, Great Smoky Mountain, AND Grand Teton National Parks, and what are the total lengths of the roads?

```
SELECT b.parkname, COUNT(*) AS road_num, SUM(r.shape_length)/1000 AS rd_length_km
FROM park_roads AS r, nps_boundary AS b
WHERE ST_Contains(b.geom, r.geom)
        AND b.parkname in('Rocky Mountain', 'Great Smoky Mountain', 'Grand Teton')
        AND rdsurface IN ('Native or Dirt', 'Gravel')
GROUP BY b.parkname;
```

This question and query were designed to locate parks with unpaved roads for ATV use and can also be modified to search for paved roads or gravel roads. From the query, it is clear that Grand Teton National Park has the most unpaved road surfaces. While some users may want to locate parks with dirt roads, others may want to avoid such roads and can choose a park that only has paved surfaces.

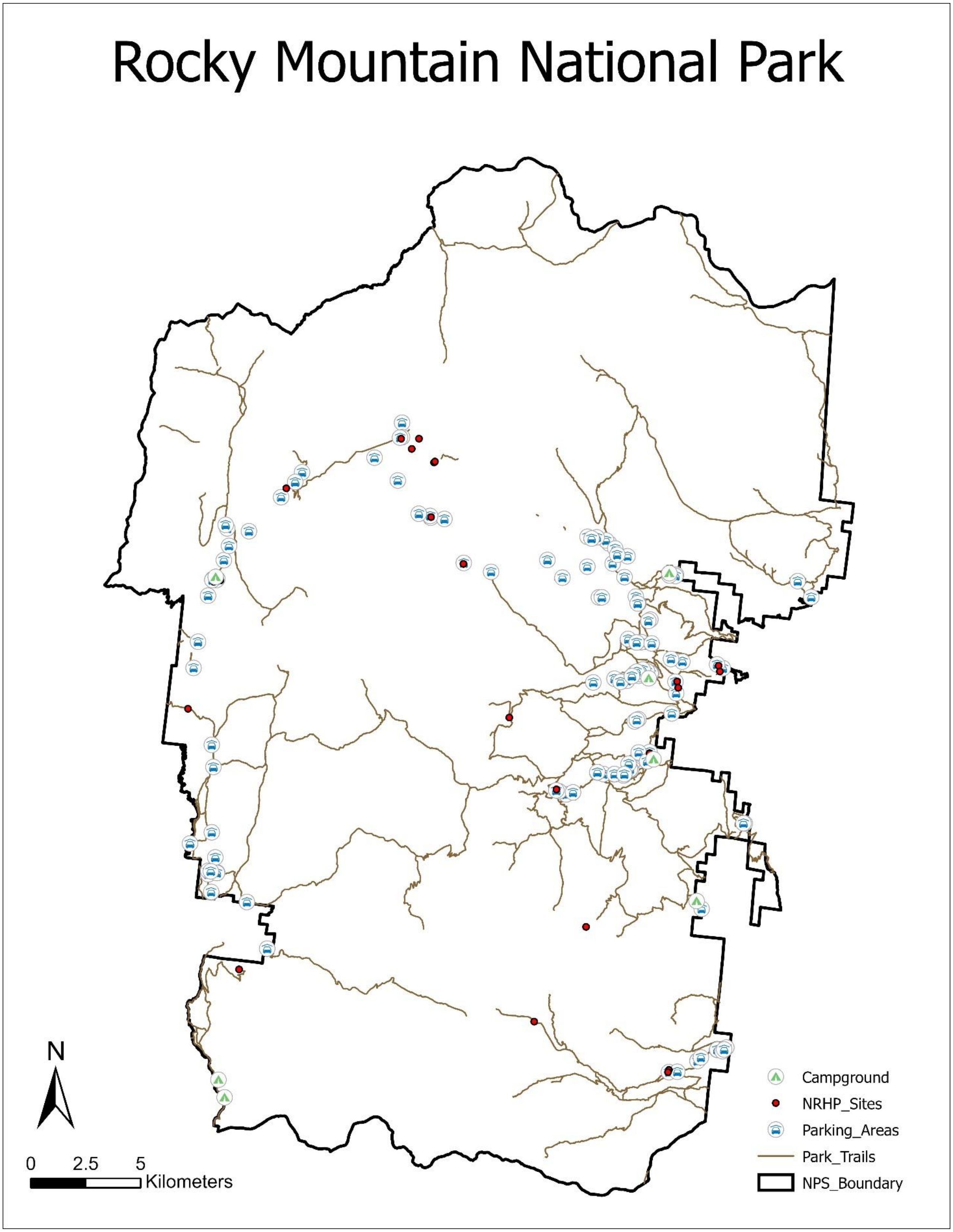

Figure 4-1. Campgrounds, historical sites, trails, and parking areas within Rocky Mountain National Park.

Question 5: Which National Park has visitation larger than the average visitation of all four National Parks in this database?

```
SELECT np_name, SUM(recreation_visitors+non_recreation_visitors) AS visitors
FROM np_visitation
GROUP BY np_name
HAVING SUM(recreation_visitors + non_recreation_visitors) > (
        SELECT AVG(s.total_visitation)
        FROM (
        SELECT np_name, SUM(recreation_visitors+non_recreation_visitors) AS total_visitation
        FROM np_visitation
GROUP BY np_name) AS s
);
```

Similar to Question 2, this question and query were designed for a user wishing to investigate and compare national park visitation data. It also demonstrates a different way to aggregate the data with a query. Again, of the four parks in the visitation data, Great Smoky Mountains has significantly more visitors than the other three national parks.

4.2 Conclusions and Future Efforts

The database was successful at returning answers to all five of the research questions posed for this project. It could potentially be useful to a user investigating the various amenities within national parks without having to look at each national park website individually or individual data downloads. However, much of the database was limited to only four national parks and some of the datasets for those parks were not entirely complete. To complete the database would require georeferencing and creating features for roads, trails, and parking areas for the remaining 419 national parks and forest reserves, as well as filing in any missing data from the nps_facilities layer. It would also be beneficial to combine visitation data for all national parks following the methodology discussed above. Completion of these tasks would be a major undertaking and was not feasible for the current undertaking, although this could potentially create a more powerful national parks database in the future.

5 References Cited

National Park Service (NPS)

2022a Office of Communications. Most Famous National Parks Set Visitation Records in 2021. https://www.nps.gov/orgs/1207/most-famous-national-parks-set-visitation-records-in-2021.htm. Accessed December 2022.

2022b Integrated Resource Management Applications Portal. Boundaries, Trails, Transportation. https://public-nps.opendata.arcgis.com/. Accessed November 2022.

2022c National Register Database and Research. Data Downloads: Spatial Data (GIS). https://www.nps.gov/subjects/nationalregister/database-research.htm. Accessed November 2022.

2022d National Park Service Visitor Use Statistics, Stats-2.9.5.21986-20221127-133810. https://irma.nps.gov/STATS/. Accessed November and December 2022.

United States Department of Agriculture (USDA)

2022 FSGeodata Clearinghouse, Download national Datasets. FSTopo RecFacility Point. https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=camp. Accessed December 2022.