

# HEAL Parks 2024 Analysis

Ronald Buie

## Table of contents

<b>1</b>	<b>changes to apply</b>	<b>3</b>
<b>2</b>	<b>Q's for Seth,</b>	<b>3</b>
2.1	Data quality . . . . .	3
2.2	Analyses . . . . .	3
<b>3</b>	<b>Front Matter</b>	<b>3</b>
3.1	Major inputs . . . . .	4
3.2	Output categories . . . . .	4
3.3	This is a quarto generated document . . . . .	4
<b>4</b>	<b>Setup &amp; Environment</b>	<b>4</b>
4.1	Secrets and tokens . . . . .	4
<b>5</b>	<b>Data Preperation</b>	<b>5</b>
5.1	Extraction . . . . .	5
5.2	Cleaning . . . . .	5
5.3	Disaggregation of pocket parks . . . . .	5
5.4	Transformation . . . . .	5
5.4.1	Date structures . . . . .	6
5.5	Create table of activities including sub areas . . . . .	6
5.6	Collapse sub areas . . . . .	6
5.7	Quality Checks . . . . .	7
5.7.1	strategy 1: correct missingness where observations are only missing but correct number of days . . . . .	7
5.8	Subset to OK data . . . . .	7
5.9	Caclulated time of day variables . . . . .	7
5.9.1	period labels in line with 2023 analysis functions . . . . .	7
5.9.2	day numbers based on sequence . . . . .	8
5.10	Study descriptors . . . . .	8

5.11	Update Study tracker . . . . .	8
<b>6</b>	<b>Load and combine prior years</b>	<b>8</b>
6.1	Import observations . . . . .	8
6.2	Import activities . . . . .	9
6.2.1	Fixing activities . . . . .	9
<b>7</b>	<b>Parks metadata</b>	<b>10</b>
7.1	Add population within half-mile radius to metadata . . . . .	10
<b>8</b>	<b>Parameterizing Data For Analysis</b>	<b>11</b>
8.1	Aggregation of periods for analysis . . . . .	11
8.2	Integrate metadata and analysis sets . . . . .	12
<b>9</b>	<b>Results</b>	<b>12</b>
9.1	General analysis notes . . . . .	12
9.1.1	Included parks . . . . .	12
9.1.2	Observation of users across target areas and time . . . . .	14
9.1.3	Parks studied in multiple years . . . . .	14
9.2	Park Utilization . . . . .	15
9.2.1	number of park users . . . . .	15
9.2.2	daily average number of park users . . . . .	15
9.2.3	daily median number of park users . . . . .	15
9.2.4	average number of park users by time period . . . . .	15
9.2.5	median number of users within each time period . . . . .	16
9.2.6	rate of daily average park use by time period . . . . .	16
9.2.7	daily average number of park users by age . . . . .	17
9.2.8	median number of park users by age . . . . .	18
9.2.9	rate of daily average park use by age . . . . .	18
9.2.10	ratio of use to half-mile catchment area . . . . .	18
9.3	Park Occupancy . . . . .	19
9.3.1	occupancy rate . . . . .	19
9.4	Activities . . . . .	19
9.4.1	number of users per activity . . . . .	19
9.4.2	rate of user activity . . . . .	19
9.4.3	rate of activity observed . . . . .	20
<b>10</b>	<b>Closure and Next Steps</b>	<b>20</b>
10.1	Data management . . . . .	20
10.1.1	Error identification and QA Work . . . . .	21
10.2	REDCap maintenance . . . . .	21
10.2.1	recommended changes to future surveys instruments . . . . .	21

10.3 Documentation . . . . .	21
10.3.1 problem notes: . . . . .	21

## 1 changes to apply

- Copy edit note: I have been trying to change the language from “target area” to “scan area”  
Q: do you want to do this now?
- Create median analyses alongside the averages  
mid process. Median not yet created for activity counts

## 2 Q’s for Seth,

### 2.1 Data quality

There are errant names in the activities list because of data entry inconsistencies in 2022. Some may involve reclassifying entries, for example, they would be “other” or fit in a pre identified category, and so look different in the observations table. Others are impossible to reclassify, such as “1”.

### 2.2 Analyses

The first analyses were named as “Averages” but were simple population counts. These have been renamed, as averages and rates are later in the analysis set. Let me know if other parameterizations are desired.

## 3 Front Matter

This document outlines procedures, technical considerations, and analytic results for the 2024 analysis of data from the HEAL Parks Study. The primary purpose of this PDF is technical review by analyst and project managers to confirm the process and data quality.

For general information about the project please review the [git](#), the project [sharepoint](#) or contact Seth Schromen-Wawrin and Ronald W. Buie

### 3.1 Major inputs

Inputs to this script are contained at `./inputs/` and include

- Access to the [ITHS REDCap](#) project: “Public Health - Seattle & King County Park Observations”, or data extracted from that project
- Final versions of non aggregated 2022 and 2023 park activities and park observations (2 files per year)
- file of park meta data, including address, zip, city, neighborhood, official name, and REDCap name for each park
- (optional) file containing record ids to redact
- (optional) file containing record ids to correct

### 3.2 Output categories

- data-metadata - tables of analyzable line item data from different stages of preparation leading up to analysis, generally in csv format
- tables - tabular outputs of analysis, generally in xlsx format
- charts - chart outputs of analysis, generally in pdf and/or png format

### 3.3 This is a quarto generated document

By rendering/knitting the qmd file, the analysis is re-executed, this document rebuilt, and new outputs are generated. To learn more about Quarto see [\[https://quarto.org\]](https://quarto.org).

## 4 Setup & Environment

This script was last executed using R version 4.4.1 (2024-06-14 ucrt).

### 4.1 Secrets and tokens

In order to pull data directly from REDCap, API information must be provided. You should create a file called “secrets.txt” in a subdirectory “./local\_only/” This file should include two lines of R code:

```
api_token    <- 'yourapikeyhere'
api_url      <- 'https://redcap.iths.org/api/'
```

Note, that the .gitignore for this project is configured to exclude your secrets.txt and anything else in the local\_only directory by default, it will not upload to github, and you will not be able to see other users' secrets.txt. They are only stored on your machine.

## 5 Data Preparation

### 5.1 Extraction

Data are extracted directly from REDCap via API.

This script requires the data and headers both be label format.

**rawOrLabel='label'**

and

**rawOrLabelHeaders='label'**

### 5.2 Cleaning

The program manager, Seth Schromen-Warren, provided corrections to data that failed QA (described below). These corrections included a list of observations requiring modification, and a list of observation that should be dropped from the data. This was based on their knowledge of the parks and study processes.

### 5.3 Disaggregation of pocket parks

Pocket parks were classified as one park (with the target areas being the pocket parks) in this year's data collection. Seth has provided a crosswalk for unpacking these into multiple parks for analysis.

### 5.4 Transformation

Data types are assigned. And character dates configured to POSIX dates.

### 5.4.1 Date structures

POSIX dates are used to generate individual variables for the day, month, and a weekend indicator variable.

Additional data corrections based on sanity checks.

Description	Details
Drop pre-study data	Study start date is 7/1/24
Drop incomplete entries	observations where the REDCap status is not “Complete”, that are missing a timestamp, or missing a park name are dropped
Drop duplicates	duplicate entries (exclusive of redcapID) are dropped. Note that where this creates incomplete days, this is corrected in later QA
Drop inaccurate subareas	Some parks were identified as having having sub area data input without having multiple sub areas (and so shouldn’t be processed as sub areas). The subarea entry for these parks is removed.
Drop observations identified by human review	Some observations were identified through review of base data by program managers. These are identified in the file “Records-to-Remove.xlsx”

## 5.5 Create table of activities including sub areas

Park activities are extracted from other park data for more accurate analysis.

## 5.6 Collapse sub areas

For this analysis, we are not using sub areas. These can be collapsed into single areas. For each observation, sub areas will be collapsed using the following rules:

- numerical observations will be added together -categorical observations become affirmative/existing if any of the subareas are affirmative/existing -timestamp of the earliest observation in the set will be used

Sub areas of the same target area will be assumed to be of the same observation period based on the following logic:

-for a sequence of sub areas observed in the same 50 hours period apparently missing sub-areas will be ignored (assumption: not all sub areas are necessary)

If non unique sub area labels are identified:

- check if the expected list of sub areas are in the data set if too many, attempt to identify redundancies and remove these if too few, interpolate missing information

In this step we also append changed record\_ids to our activity table as “record\_id\_aggregated” and save the updated activity table.

## 5.7 Quality Checks

The QA process attempts to identify and correct errors. The process initially performs a series of checks on all park data and reports results. For each park that fails QA, various strategies are executed to attempt to cure that park’s data. The final results, including which strategies were executed, and the final QA status are saved in a csv file for review and manual correction where necessary.

### 5.7.1 strategy 1: correct missingness where observations are only missing but correct number of days

This strategy looks at the number of days of data observed, and, if only 3 days are observed, then checks to confirm that, for each target area, 8 or fewer observations are made. If both of these conditions are met, this strategy attempts to insert the missing observations as blank entries in the parts of the day that appear to be missing them.

## 5.8 Subset to OK data

The below steps assume that complete data are being used. Data that did not pass QA are not utilized in the rest of the preparation or analysis.

## 5.9 Calculated time of day variables

(move this up after fixing code to work earlier)

### 5.9.1 period labels in line with 2023 analysis functions

time of day labels are created based on the available data. In 2024 we requested observers to specify their reporting period, so this should be trivial for non corrupted entries. We recreate the variable “period based on sequence” to reduce work in refactoring last year’s code.

### 5.9.2 day numbers based on sequence

add a counter for 1st through 3rd day of observation.

## 5.10 Study descriptors

A descriptor is added to this year's results:

Variable	Value
StartDate	"2024-07-08"
StudyDescription	"2024 annual study"
study_count	"1" (this will be adjusted in later steps)

## 5.11 Update Study tracker

For the 2023 analysis we advised creating a study tracker. For each time a park is studied a new entry is created. Each entry should include the park name, the date the study started, and a brief description of the study.

In a later step we will use this table to update the study count to our parks

# 6 Load and combine prior years

Going forward, we intend to conduct analyses over all available years of data. Previous year's data is provided in the inputs folder. Future executions should only need the last and the current years data, as the saved tables will include all prior data.

## 6.1 Import observations

For 2022 data there are 6 observation days per park. These have been parameterized as 2 studies of 3 days each.

per Seth 4/25/2024:

When we did the observations in 2022, we did two different time periods: early summer and late summer. Both are of-quality. One option would be to combine the 6 observation days for each park as 2022 data (like you said). Another option would be to have two different 2022 data sets. If you have a preference, go with that.



From my end, I lean towards combining them into one 2022 analysis with 6 observation days, unless it makes issues in the future to have some time periods with 3 observation days and some have 6. My feeling is having one 2022 dataset is the easiest way to communicate it. Keeping it separate does not add anything particularly unique, like pre/post opportunities for a park construction project.

## **6.2 Import activities**

### **6.2.1 Fixing activities**

Several activities can be reclassified for more accurate analysis. This is a manual step, so information is provided below to assist.

Here is a list of activities presented to observers in 2023 and 2024. They may also enter “other” and type a custom entry.

Baseball/softball, Basketball, Bike Riding, Catch (any sport), Climbing, Dance/Aerobics (dance/step aerobics), Fitness stations, Football, Frisbee, Jumping (rope, hop scotch), Lacrosse, Lying down, Picnic (food involved), Playground activities, Running, Sitting, Skating/skateboarding, Soccer, Standing, Strength exercises (pull ups), Tag/chasing games, Tennis/racquetball, Volleyball, Walking, Other

This year we focused on fixing errors in 2022 activities to help them align to this list.

We also identified one 2024 activity where the observer entered “other” but did not enter something into the box. this has been changed to “other”.

Procedure:

For 2022 activities

Where an activity was not classified as “other” AND appears to match, but not precisely conform to, the 2024 activities list, update both the activity table and the observations table, keeping the activity in its observation step (primary, secondary, or tertiary activity).

If the activity was not classified as “other” AND does NOT appear to match the 2024 activities list, then, in the observations table, change the activity to “other” and move the unmatched name to “other description” of the appropriate observation step. Leave unchanged in the activity table

After making corrections, here are activities in our data that are not in the above list. These should be manually reviewed for possible further correction.

Swinging, Bike riding, Pickle ball , Fishing, Fishing , Construction , Dance/aerobics (dance/step aerobics), Riding a scooter , Scooters, Planting, Gardening, Stroller, Kite flying , Grilling, Rolling in grass, Driving, Handball, Hand ball, Water activities, Concert in the park , Concert in the park, Water activities , Cooking, Picking up trash, Flying kite, Wheelchair

, Carpentry, Gardening, Gardening , Gathering of people doing various stuff.. sports, eating and dancing , Organized events - various sport, food and activities , Group event of f sports, food, dance and running around , Sorting produce , Lawn mowing , Lawn mowing, Organized play at the park and drink and dance, Musical in the play in city of burien town square park , Musical play in city of burien town square park , Mowing grass, Dancing, Motor dirt bike, Dirt bike riding, Driving , Pathway construction , Leafblowing, Pressure washing on crane, Paddle boarding , Games, Kids plying, Playing, Supervise, Pushing kid, Wlking dog, Play area, Stretching, Wking out, Tennis, Watching, Kck ball, Spectator, Throw/ball, Wthg game, Skateboardg, Wtchg, Exercise, Napping, Reading, Laying down, Jump rope, Exercising, Jogging, Sleeping, Play/dog, Bbq, Yoga, Plyg w/dog, Skateboard, Kickball, Riding bikes, Running tag, Biking, Skating, Picknicking, Catch, Park empl, Parent/supervise, Garden wk, Picking, B-ball, V-ball, Picnic, Laying dwn, W-out, Jumping rope, Runnng, Laying, Supervising, Jumping, Sliding, Jogger, Riding, Runing, Slide, swing, climb, Lying, Picnicking, Skateboarding, Tag, Horseshoe, Horseshoes, Horse shoes, Bicking, Ring toss, Picnicing, Kick ball, Rollerblading, Park bench, Playground, Picnic shelter, Grass play area, Ticket stand, Pathway, Teen garden, Grass, Playing swimming, Illegal activity , King county worker changing trash cans, Pickle ball, Worker cutting grass, Worker open all gates, Worker grass, 2 staff pushing stadiums seats to place on field, Worker driving a cart, Workers setting up, Staff cleaning , Woman stretching , Drills on sidewalk, Working, Working snack stand, Swimming, Paddleboarding, Standing in the water, Kayaking, Playing in the water, Sleeping in tents , Sleeping in tent , Working on construction , Construction, Working on excavator , Sleeping in sleeping bag , Sleeping , Working , Playing in city field water sprinkler, Playing in the sand, Staff cleaning empty trash, Kids playing in waterfall area, Staff cleaning from chase, Staff cleaning skate area, Boating, Riding segways, Park staff cleaner, Waterfall , Soccer volleyball , Sleeping under sleeping bag , Sleeping in sleeping bag, Children set up cones for exercise., Packing bikes, Play golfing, Park cleaner, Driving county car, Setting up chairs, Breaking down chairs, Frisbee

## 7 Parks metadata

Metadata are provided for each park by Seth. The formal name, address, city, zip, neighborhood, tract, equity score, longitude, latitude, image status, planned park change notes, and general notes for each park will be appended to the analysis table.

The following parks (if any) are missing from the Parks Master sheet:

The following parks (if any) are missing Longitude and/or Latitude:

### 7.1 Add population within half-mile radius to metadata

For each park we estimate the number of people within a half-mile. This is based on the generally accepted threshold of a [10-minute walk](#) to a park as a metric goal.

This relies on access to APDE population data and use of the RADS library and on park longitudes and latitudes provided in the meta data.

## 8 Parameterizing Data For Analysis

### 8.1 Aggregation of periods for analysis

For all analyses, we do not want to distinguish between the first and second of a period (e.g. morning1 and morning2). We aggregate these according to the following:

positive indicators	aggregate to
accessible	Yes if any Yes
usable	Yes if any Yes
lit	Yes if any Yes
occupied	Yes if any Yes
supervised	Yes if any Yes
organized	Yes if any Yes
equipped	Yes if any Yes

  

counts of	are aggregated as
num_child_prim	ceiling of mean
num_child_snd	ceiling of mean
num_child_tert	ceiling of mean
num_child_quat	ceiling of mean
num_teen_prim	ceiling of mean
num_teen_snd	ceiling of mean
num_teen_tert	ceiling of mean
num_teen_quat	ceiling of mean
num_adult_prim	ceiling of mean
num_adult_snd	ceiling of mean
num_adult_tert	ceiling of mean
num_adult_quat	ceiling of mean
num_senior_prim	ceiling of mean
num_senior_snd	ceiling of mean
num_senior_tert	ceiling of mean
num_senior_quat	ceiling of mean

Activities are independently captured in the activity table and not aggregated here. The non aggregated observations table may contain multiple activities separated by a “;” where they

were grouped from sub-areas. It is recommended to use the activity table for analysis, and join other park observation data to it using the record\_id and record\_id\_aggregated where necessary.

## 8.2 Integrate metadata and analysis sets

We join our metadata and park observations for analysis and for use in other software such as excel.

## 9 Results

Results are saved in multiple locations:

- an excel workbook that contains a separate page for each table of analysis results
- folders with charts and tables of results designed to be integrated into documents
- outputs of various steps of the metadata, QA, and final analysis ready sets

### 9.1 General analysis notes

#### 9.1.1 Included parks

Table 5: Parks Included In Analysis

Park Name
Annex Park
Arbor Lake Park
Beverly Park Elementary School
Bicentennial Park
Boulevard Lane Park
Brighton Playfield
Cascade View Community Park
Cecil Memorial Park
Cedarhurst Elementary School
Chelsea Park
Crestview Park
Crystal Springs Park
Dick Thurnau Memorial Park
Dotty Harper Park
Duwamish Gardens

Table 5: Parks Included In Analysis (*continued*)

Park Name
Duwamish Hill Preserve
Duwamish Park
Five Mile Lake Park
Fort Dent Park (North)
Fort Dent Park (South)
Garfield Playfield
Greenbridge - Community Garden
Greenbridge - North Playground
Greenbridge - Plaza
Greenbridge - W4 Playground
Greenbridge - Wave Playground
Greenbridge - West Playground
Gregory Heights Elementary School
Hazel Valley Elementary School
Hazel Valley Park
Hazelnut Park
Hilltop Park
Jacob Ambaum Park
Joseph Foster Memorial Park
Lake Burien School Memorial Park
Lake Geneva Park
Lakeview Park
Linde Hill Park
Manhattan Park
Maple Valley Heights Park
Maplewood Park
Marra Desimone Park
Mathison Park
Midway Park
Moshier Memorial Park
Mount View Elementary School
North Shorewood Park
Puget Sound Park
Riverton Park
Roxhill Park
Salmon Creek Park

Table 5: Parks Included In Analysis (*continued*)

Park Name
Seahurst Park
Seola Gardens - North Playground
Seola Gardens - Park
Seola Gardens - South Playground
Shorewood Elementary School
Skyway Park
Soos Creek Park (Gary Grant)
South County Ballfields
Southern Heights Park
Steel Lake Annex Park
Steve Cox Playfield
Town Square Park
Tukwila Community Center
Tukwila Park
Tukwila Pond
White Center Heights Park

### 9.1.2 Observation of users across target areas and time

Several of the metrics rely on counts of people observed. The underlying data are of people observed within an observation period (e.g. “morning1”) and target area. It is possible that the same people may be observed across multiple blocks of time and multiple target areas.

Because of this user counts are more accurately understood as “person time of use per target area” and represents a target area being used by a person within the observation period. This explanation accounts for people being counted multiple times by crossing target areas during the observation period.

### 9.1.3 Parks studied in multiple years

The below analyses include parks studied from 2022 to present unless otherwise indicated. For these analyses all available data are used. This means that if the same park has been studied for multiple years, all years are included.

## 9.2 Park Utilization

### 9.2.1 number of park users

The total number of users observed for a given park across all available data.

Notes:

- Due to the study design, user counts are subject to both over and under counting
- Does not adjust for number of studies or days of park observation. Therefore more heavily studied parks will tend towards higher counts.

### 9.2.2 daily average number of park users

The average number of users observed in the park per day across all available data.

### 9.2.3 daily median number of park users

The median number of users observed in the park in a given day across all available data.

### 9.2.4 average number of park users by time period

The average number of users within each time period.

For each park:

$$(\text{Average Park Users By Period}) = \frac{\sum_{d=1}^3 (\text{People}_p)_d}{3 * (\text{number of times park studied})}$$

where  $_d$  is the day of study and  $(\text{People}_p)$  are the sum of number of people observed in a time period of a given day.

The people within a given time period and day is defined as the average (rounded up) of people observed in a time period of a given day and target area  $_t$ , summed across across all target areas. This is precalculated in the SOPARCAggregated table.

$$\text{People}_p = \sum_{t=1}^n \left( \left\lceil \frac{(\text{people observed first half of period}) + (\text{people observed second half of period})}{2} \right\rceil \right)_t$$

Notes:

- User counts are more accurately understood as “person time of use per target area”
  - “Person using target area during the observation”
- If taken strictly as “people using a park” then this may be an over or under count
  - If in the first morning observation 2 people are observed, and the second 3, this may be 3 unique people, or as many as 5, but we calculate this as 2.5 and round up to 3.
  - If two people walk across all target areas during an observation period, they would be counted each time. With 10 target areas, this would be 20 people observed.

### 9.2.5 median number of users within each time period

The median people within a given time period and day is defined as the median number of people observed in a time period across all days of a given park.

$$(\text{Median Park Users By Period}) = m(\sum_{d=1}^3 (\text{People}_p)_d)$$

where  $_d$  is the day of study and  $(\text{People}_p)$  are the sum of number of people observed in a time period of a given day.

Notes:

- User counts are more accurately understood as “person time of use per target area”
  - “Person using target area during the observation”
- If taken strictly as “people using a park” then this may be an over or under count
  - If in the first morning observation 2 people are observed, and the second 3, this may be 3 unique people, or as many as 5, but we calculate this as 2.5 and round up to 3.
  - If two people walk across all target areas during an observation period, they would be counted each time. With 10 target areas, this would be 20 people observed.

### 9.2.6 rate of daily average park use by time period

The proportion of the total users observed within each time period.

For each park:



$$(\text{Rate of Park Use By Period}) = \frac{\sum_{d=1}^3 (\text{People}_p)_d}{\sum_{d=1}^3 \text{People}_d}$$

where  $_d$  is the day of study and  $(\text{People}_p)$  are the sum of number of people observed in a time period of a given day.

The people within a given time period and day is defined as the average (rounded up) of people observed in a time period of a given day and target area  $_t$ , summed across across all target areas. This is precalculated in the SOPARCAggregated table.

$$\text{People}_p = \sum_{t=1}^n \left( \left\lceil \frac{(\text{people observed first half of period}) + (\text{people observed second half of period})}{2} \right\rceil \right)_t$$

The total number of people in a day,  $\text{People}_d$ , is defined as the sum of all  $\text{People}_p$  within a day.

Notes:

- This is accurately understood as the rate of park use during the period
- There is likely still some error from the under and overcounting of the underlying counts, but the more true that over and under counting is randomly distributed across all time periods, the less true this is.
- Rate is within-park (each park totals to 100%)

### 9.2.7 daily average number of park users by age

The daily average number of users observed in each age group.

Computationally this measure is similar to the average users by time period above.

Notes:

- User counts are more accurately understood as “person time of use per target area”
- If taken strictly as “people using a park” then this may be an over or under count
  - If in the first morning observation 2 people are observed, and the second 3, this may be 3 unique people, or as many as 5, but we calculate this as 2.5 and round up to 3.
  - If two people walk across all target areas during an observation period, they would be counted each time. With 10 target areas, this would be 20 people observed.
- “Teens” were not a category in 2022, meaning all apparant minors are classified as “child”. Teens may be underweighted for affected parks.

### 9.2.8 median number of park users by age

The median number of users observed in each age group.

Computationally this measure is similar to the median users by time period above.

Notes:

- User counts are more accurately understood as “person time of use per target area”
- If taken strictly as “people using a park” then this may be an over or under count
  - If in the first morning observation 2 people are observed, and the second 3, this may be 3 unique people, or as many as 5, but we calculate this as 2.5 and round up to 3.
  - If two people walk across all target areas during an observation period, they would be counted each time. With 10 target areas, this would be 20 people observed.
- “Teens” were not a category in 2022, meaning all apparent minors are classified as “child”. Teens may be under weighted for affected parks.

### 9.2.9 rate of daily average park use by age

The proportion of the total users observed within each age group.

Notes:

- This is accurately understood as the rate of park use by each age group
- There is likely still some error from the under and over counting of the underlying counts, but the more true that over and under counting is randomly distributed across all time periods, the less true this is.
- Rate is within-park (each park totals to 100%)

### 9.2.10 ratio of use to half-mile catchment area

The ratio of how many users were observed per day on average relative to how many people live within 0.5 miles of the park.

This is provided per 1,000 residents to improve readability.

Notes:

- The ratios are calculated on non-rounded data, and then rounded. The average number of users and populations provided in the formatted word document are rounded. This causes a rounding error where you wouldn’t get the exact ratio if you were to divide these users and populations. For accurate results use the numbers provided in the pivot ready tables.

## **9.3 Park Occupancy**

### **9.3.1 occupancy rate**

The percentage of observations where at least one user was observed in the park.

For each park, the number of observation periods with any target area in occupied status is divided by the total number of observation periods (24)

Notes:

## **9.4 Activities**

### **9.4.1 number of users per activity**

The number of users ever observed doing the activity in the park.

Notes:

- Note, includes ALL years. Due to the varying number of times a park has been observed and is not adjusted for how many times a park has been observed
- user counts are subject to over and under counting due to how observations were conducted and aggregated
- If an individual is observed doing a different activity in the second half of a quarter than the first half, these will be counted as distinct activities, and so not averaged. This can result in slightly higher counts of people engaged in activities than the average user counts.

### **9.4.2 rate of user activity**

The percentage users observed doing the activity in the park.

For each park, the number of users engaged in an activity is divided by the total number of users observed in the park throughout all days and studies observed.

Notes:

- All activities listed have at least 1 participant, but some may show 0 in the prepared tables due to being less than 0.01%
- The total rate of all user activities listed will equal 100%, the total amount of activity observed.
- This is calculated on the non aggregated population counts.

### 9.4.3 rate of activity observed

The percentage of observations where at least one user was observed doing the activity in the park.

For each park, the total number of periods the activity was observed is divided by 24 \* (the number of times studied), which is the total number of observations periods possible for any particular activity.

Notes:

- These rates are mostly independent of each other and do not have an additive meaning. This is because multiple activities may be observed in a single observation period. E.g. “walking” may be observed in all 24 periods, and so have an observation rate of 100%, and “sitting” may be observed in 6 periods and so have a rate of 25% for the same park.

## 10 Closure and Next Steps

### 10.1 Data management

Files in the provided data-metadata directory should be retained. These include QA information, updated metadata, and observational data at different stages of data preparation.

Most critical is the maintenance of the version of the data used for the above analysis. Over the years, these would be the ones expected for use in further analytic work and for comparing year-over-year results.

The analysis-ready files include:

- SOPARCAAnalysisSetAllYears.csv (fully cleaned park observation)
- SOPARCAAnalysisSetAggregatedPeriodsAllYears.csv (analysis set with observations aggregated by period, the level of analysis most commonly used)
- SOPARCAActivities.csv & SOPARCAActivitiesExpanded.csv (all activity data at the individual record level. The expanded version includes some park information already attached for human readability)

The contents of this working directory (this script, the compiled pdf report, the inputs folder, and the outputs folder) should be copied to this project’s [sharepoint folder](#)

### 10.1.1 Error identification and QA Work

- Fix in-report print of results to just show 1-3 as a sample
  - fix broken results prints
- The sub area aggregation code was not updated this year because no sub areas were observed.

## 10.2 REDCap maintenance

Following closure of data collection the REDCap project should be moved into “analysis/cleanup”. It is recommended to keep it in this stage until finished with any future studies and analyses in this line of work, e.g. if planning to use the data from this study in the future, it is good to keep this REDCap project in analysis mode for review. There is an additional “completed” status, which should generally only be used when completely finished with the body of work. This status makes the project largely inaccessible. Notably, changing the project to “complete” status would also break this script (or require it to be fed a different data source rather than pulling from the REDCap API.)

I recommend conducting future data collection efforts in their own REDCap projects, naming them similarly, such as “Public Health - Seattle & King County Park Observations: title\_of\_study” The project used for this analysis has been renamed to: Public Health - Seattle & King County Park Observations: 2024 Annual Study

### 10.2.1 recommended changes to future surveys instruments

- Include additional testing and possibly consultation with ITHS REDCap staff for issues with data upload from cell phones. This cause problems for observers.
- A participant was able to choose “other” activity but not enter something in the other description. This shouldn’t have been possible. Fix the instrument validation to prevent this. (2024 study, record\_id 15651 in raw, 9553 in prepared)

## 10.3 Documentation

Documentation for this project is available in [git hub](#) and in the [project sharepoint library](#).

### 10.3.1 problem notes:

- Latex math chunks do not render correctly when the README.me is viewed on github
- Latex in-line math does not render at all when the README.me is viewed on github
- Some package/visualization is preventing a non-html compilation of github markdown