

Using Open Data to Model Recreation in Seattle's Urban Parks

Capstone Project

University of Washington's
Professional Master's Program
in Geographic Information Systems
for Sustainability Management

In Cooperation with Natural Capital Project
Geography 569 GIS Workshop
August 21, 2015

By Jeremy Forst and Malena Foster

Table of Contents

1	Recommended Course of Action	1
2	Introduction.....	1
3	Design and Methods.....	4
	3.1 Validating Flickr and Twitter Data against Tacoma Visitation Data:	6
	3.2 Seattle Parks Data Analysis.....	8
4	Results	13
	4.1 Tacoma Parks Validation	13
	4.2 Seattle parks.....	14
5	Discussion	22
	5.1 Tacoma Validation Discussion	22
	5.2 Seattle Park Feature Analysis	25
6	Business Case and Further Recommendations	26
7	Literature Cited.....	27
8	Data Links	27
	Appendix A: Scripts	
	Appendix B: EarthCorps Habitat and NOAA C-CAP Aggregations	

List of Figures

Figure 1. Comparison of PUD and Empirical visitation data across different worldwide locations.	2
Figure 2. Comparison of developed and developing sites (Wood et al, 2013).	3
Figure 3. Comparison of Natural and Cultural Recreational Sites (Wood et al, 2013).	3
Figure 4. Comparison of datasets across countries	3
Figure 5. Empirical data versus Flickr photo user days for Iowa and Minnesota State Parks (Keeler et al, 2015).	4
Figure 6. Project Workflow Diagram.....	5
Figure 7. Parks across the City of Seattle.	9
Figure 8. Scatter plots of park visitation against PUD and TUD (Point Defiance Park omitted).....	13
Figure 9. (a) Trendlines for August TUD against visitation and (b)the same data with the outlier removed.....	14
Figure 10. Results of comparison of TUD, PUD, and Shape Area data.	14
Figure 11. Results of comparisons between TUD, PUD, and Habitat/Land Cover data.....	15
Figure 12. Results of comparison between TUD, PUD, and demographic variables.	16
Figure 13. Results of comparison between TUD, PUD, and Accessibility variables.....	17
Figure 14. Distribution of Average Annual Photo User Days (Flickr) in Seattle Parks	20
Figure 15. Distribution of Average Annual Twitter User Days in Seattle Parks	21
Figure 16. Wright Park Arboretum and Conservatory	23
Figure 17. Meadow Park Golf Course	24
Figure 18. Ruston Way Escarpment. Ruston Way Park is one of a series of parks along the walking and biking trail that follows the coastline.	24

List of Tables

Table 1. Relevant Field names in Tacoma Parks Visitation Study (Herbert 2010).....	6
Table 2. Original Files used in Flickr/Twitter data validation in Tacoma	7
Table 3. Sources of Data Acquired.....	10
Table 4. Proposed predictor variables grouped by category.....	11
Table 5. Comparison of samples means of Flickr PUD where an attribute is present or not using a Mann Whitney U t-test.....	18
Table 6. Comparison of samples means of Twitter User Days (TUD) where an attribute is present or not using a Mann Whitney U t-test.....	19

List of Abbreviations

TUD – Twitter User Days

PUD – Flickr Photo User Days

NatCap – The Natural Capital Project

InVEST – Integrated Valuation of Ecosystem Services and Tradeoffs. NatCap’s ecosystem service valuation software.

1 RECOMMENDED COURSE OF ACTION

Based on Twitter User Days (TUDs) and Flickr Photo User Days (PUDs), the factors that appear to draw the most people to City of Seattle parks are park size, presence of water bodies, population density in the surrounding neighborhood, number of nearby bus stops, and a number of facilities attributes including beaches, Fishing, Off-leash dog areas and picnic sites. TUDs and PUDs have been shown to be reasonable proxies for actual visitation data in Tacoma when outliers are eliminated from the analysis. Our findings also add to the growing body of evidence showing the use of Twitter User Days to be a reasonable proxy for parks visitation data by showing a strong correlation with visitation in the city of Tacoma.

Therefore we recommend that NatCap aggregate the findings of this report into the growing set of similar studies in other locations and use the results to refine and add functionality to the InVEST recreation model.

We also recommended that NatCap explore ways to model other ecosystem services of urban parks, such as air filtering, hosting urban biodiversity, and measurable health benefits to urban populations.

2 INTRODUCTION

The Natural Capital Project (NatCap) is looking to identify what park attributes lead to higher visitation rates in urban parks. Urban parks are an important integration in urban life but research into what makes one park more valuable than another is limited due to the difficulty in obtaining visitation data. In order to investigate this question, the InVEST Recreation model has been created to extract Flickr data users as proxy for actual visitation data within parks. These variables can be used to investigate which park attributes contribute to higher visitation numbers. The Natural Capital Project is searching to investigate the use of this model in Seattle urban parks and validate the Flickr data against actual visitation data.

Background

Empirical visitation data is expensive to collect and difficult to come by. The Natural Capital Project has been looking for a proxy to visitation by using social media. Flickr data was chosen because the data set is globally extensive, geotagged, has been available since 2005, and is continually expanding. The Flickr application program interface (API) also facilitates extraction easily. For this reason, it was one of the primary social media applications considered for the InVEST recreation model (Wood et al, 2013).

Previous research by the Natural Capital Project has looked at the use of Flickr data at 836 recreational sites around the world, ranging from parks to amusement and cultural attractions and compared this to empirical visitation data. The Flickr data was used to compute an average annual Photo User Day (PUD) at each site location from 2005 to 2012. The average annual user day takes into account the geotagged location of a visitor but counts all photos taken in a single day as one user day. Photos taken across multiple days are considered separate events. And average of these events over the course of multiple years produces the PUD value for each location. A similar value is determined based on empirical data from each site. The Natural Capital Project was able to show that there was a relationship between the empirical data and the proxy values at all locations as shown in **Figure 1** below (Wood et al, 2013).

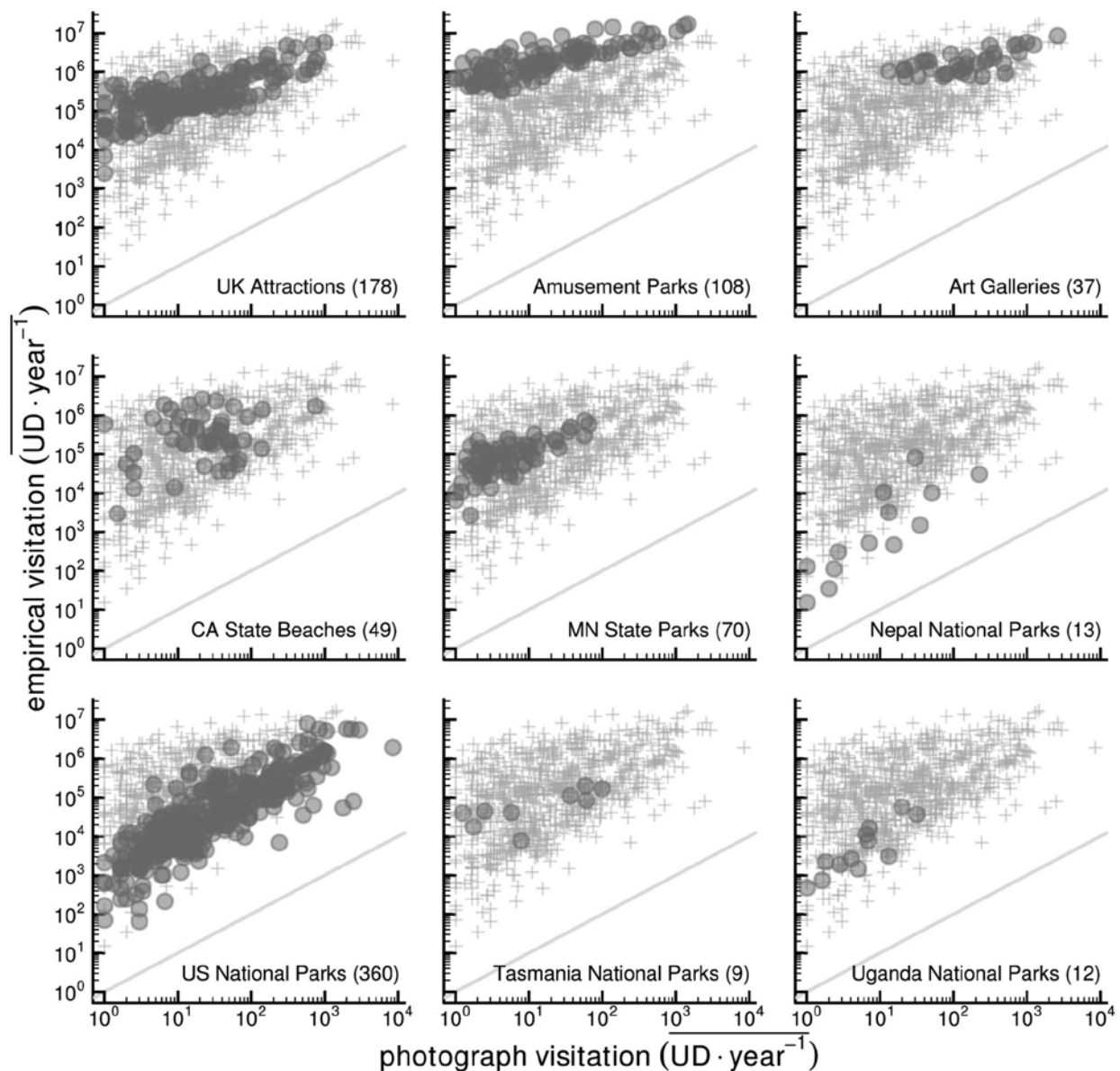


Figure 1. Comparison of PUD and Empirical visitation data across different worldwide locations.

The study indicates that the type of attraction and income-level of a country do not change the observed relationship between the empirical user day (EUD) and PUD (see **Figure 2**). However, socioeconomic conditions do explain variation across the sites in the number of photos taken and empirical visitation accounts as shown in **Figure 3**. Despite this, the slope on average remained at 0.70 with absolute visitation varying across sites. The study indicated that the proxy measurement of Flickr Photo User Days can be applied almost anywhere as shown in **Figure 4**. This relationship was consistent across developing and developed countries, areas where there was little data versus a great deal, and urban areas versus wilderness sites (Wood et al, 2013).

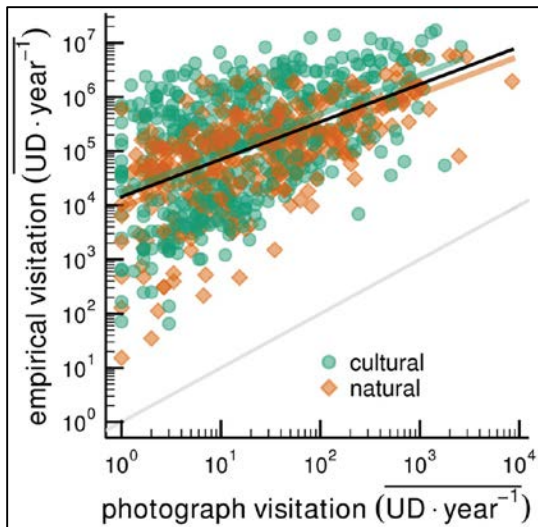


Figure 3. Comparison of Natural and Cultural Recreational Sites (Wood et al, 2013).

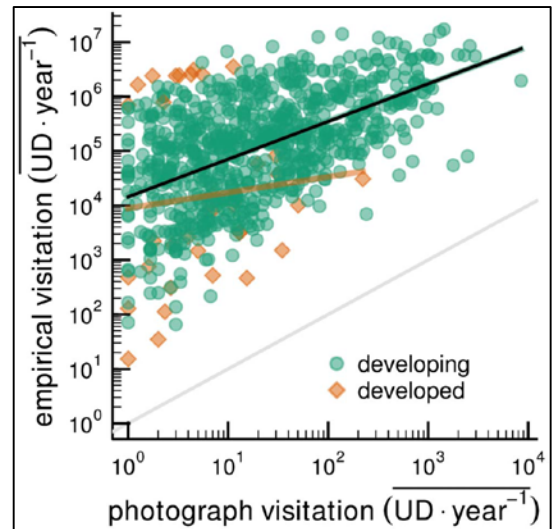


Figure 2. Comparison of developed and developing sites (Wood et al, 2013).

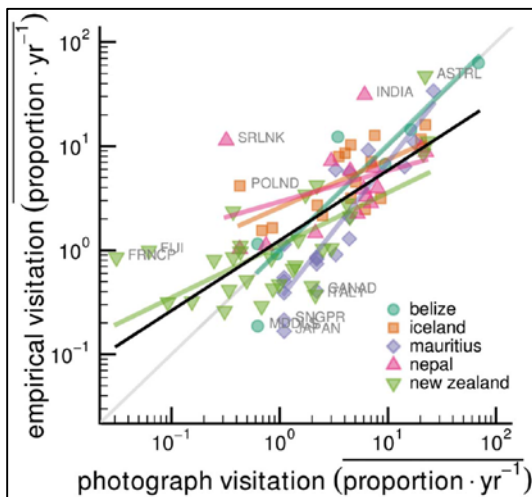


Figure 4. Comparison of datasets across countries (Wood et al, 2013).

Based on the promising data from this study, another one was conducted on Iowa lakes and Minnesota state parks as seen in **Figure 5**. This study showed not only a strong correlation between empirical visitation trips and PUD but that there were some variables predictors for an increase in visitation at these locations. The highest predictor variables were lake size, water clarity, near-lake population, presence of a boat ramp, and the state (Iowa or Minnesota) (Keeler et al, 2015).

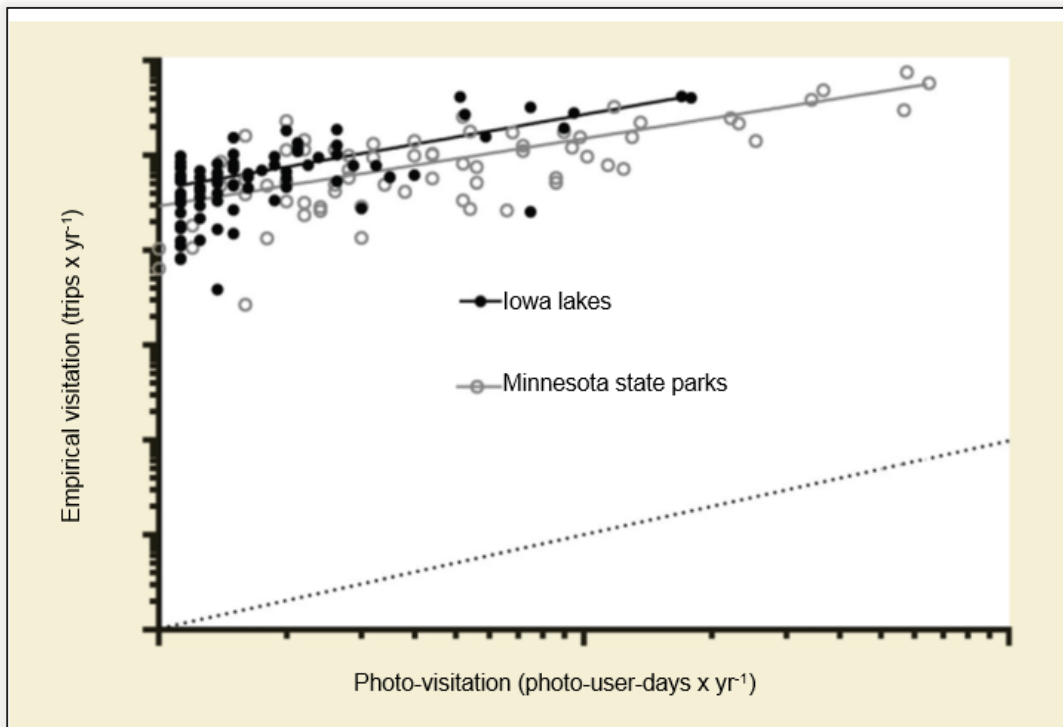


Figure 5. Empirical data versus Flickr photo user days for Iowa and Minnesota State Parks (Keeler et al, 2015).

From the promising previous studies, the Natural Capital Project is looking to determine whether or not Flickr user days can be used to study what park features attract higher visitation rates. The current InVEST recreation model uses Flickr to extract photo user days at a location of interest and the model is used as a proxy for visitation data. The Natural Capital Project has conducted previous research on New York and the Minneapolis-St. Paul area and is looking to include Seattle parks in their dataset.

The Natural Capital Project is also planning a next version of the InVEST recreation model that will include the option of also using Twitter to extract Twitter user days at locations of interest.

3 DESIGN AND METHODS

The examination of the Seattle parks data was a two part question: Determining if the Flickr output from the InVEST model could be used as a proxy for visitation data and using the Flickr PUD output data to examine any significant predictor variables. A workflow diagram is shown in **Figure 6** describing the process.

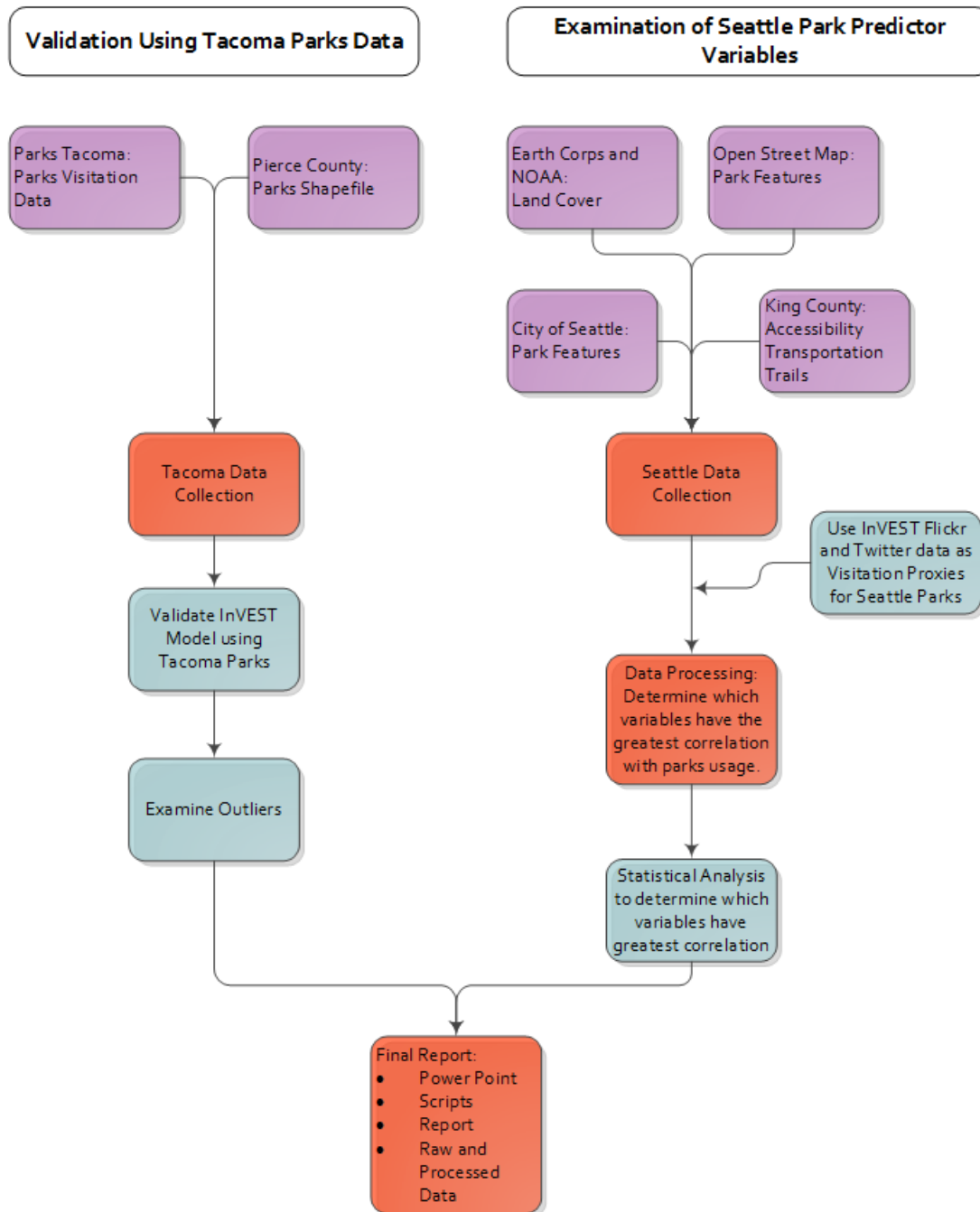


Figure 6. Project Workflow Diagram

Because no visitation data was found for the City of Seattle, a study done on Tacoma (Hebert, 2010) was used as representation for Seattle for the validation part of the project. Tacoma parks data was collected along with Twitter and Flickr data discussed in Section 3.1 below. The second part of the project collected data for Seattle parks to compare to the outputs from the InVEST model and analysis was done as shown in Section 3.2.

3.1 VALIDATING FLICKR AND TWITTER DATA AGAINST TACOMA VISITATION DATA

Since we are using Flickr data as a proxy for visitation in Seattle, we need to be sure that it is an accurate measure of parks visitation in Seattle. Since there is no park visitation data for City of Seattle parks, we were required to look for a substitute city in which to perform the validation. The city of Tacoma was selected for two reasons: It has similar demographics and park characteristics to Seattle, and secondly, a parks visitation report, conducted in August, 2009, exists in Tacoma. The report (Hebert 2010), commissioned by Metro Parks Tacoma and conducted by Hebert Research, consists of an intercept survey distributed at 16 representative parks within the city, totaling 769 samples.

Table 1. Relevant Field names in Tacoma Parks Visitation Study (Herbert 2010).

Variable	Description
GrpSizeRes	Estimated average size of groups - Residents of Tacoma
GrpSizeOut	Estimated average size of groups - Outside Tacoma
GroupsRes	Estimated Annual Groups of Visitors - Tacoma Residents
GroupsOut	Estimated Annual Groups of Visitors - Outside Tacoma
TotalRes	Estimated Annual Visitors = Tacoma Residents
TotalOut	Estimated Annual Visitors - Outside Tacoma
TotalVis	Estimated Total Annual Visitors

We discovered that the current InVEST recreation model can only produce Flickr PUD results for parks larger than 40,592 square meters (approximately 10 acres). The Natural Capital Project helped us work around this by providing raw Flickr and Twitter data that are not available in the current version of the InVEST recreation model so that both datasets could be compared. Dave Fisher provided us a shape file with 2012-2014 Twitter User Days (TUD) for all parks in King County and Pierce Counties, along with additional data tables with 2005-2014 Flickr data, plus tables with mean TUDs for the month of August as well as PUDs for August 2009 (the month the Hebert survey was conducted) for the same area. All four files contained a unique Park ID field that was used to join them in ArcMAP. In order to complete the preparation for analysis, the actual visitation (Hebert) data needed to be added to the table. The park name field was used for this join. Because the names of some of the parks differed between the Hebert data and park names in the Pierce County parks shapefile some name modification was needed to make the records match.

Table 2. Original Files used in Flickr/Twitter data validation in Tacoma

File Name	Type	Description	Area	Field to join with TUD shapefile
Tacoma_Parks_Visitation_Data-Hebert_Report.csv	CSV	Visitation data adapted from Metro Parks Tacoma report (Hebert 2010)	Tacoma	Park_Name
Tacoma_KingCo_parks_merged_TUD.shp	Shapefile	All twitter user days for 2012-2014	King and Pierce Counties	PID
Tacoma_KingCo_parks_annual05-14_PUD.csv	CSV	All Flickr photo user days for 2005-2014	King and Pierce Counties	PID
Tacoma_KingCo_parks_aug09pPUD.csv	CSV	Flickr photo user days for the month of August, 2009. Corresponds to the time period of the Hebert survey	King and Pierce Counties	PID
Tacoma_KingCo_parks_augTUD.csv	CSV	Twitter user days for the month of August in 2012, 2013 and 2014	King and Pierce Counties	PID

Two of the parks in the Hebert report, Fort Nisqually and Brown’s Point playfield could not be used in the analysis because they did not appear in the Pierce County parks shapefile.

The records were combined and exported as a new, unified shapefile called Tacoma_PUD_TUD. It became clear that Point Defiance park was an extreme outlier, because of its very high rate of visitation (nearly 10 times that of Ruston Way, the second most popular park) and was therefore removed in order to better illustrate the relationship of the variables. The data from this shapefile was exported to Excel, all unnecessary fields were removed, and scatterplots were produced for each of the dependent variables (Figure 8)

Trend line diagrams (Figure 9) were produced to show relationships between visitation and TUDs both with and without further outliers removed. R Square values were calculated to quantify how well TUDs predict visitation.

Lastly, we examined the characteristics of outliers – parks whose visitation did not appear to be effectively predicted by PUDs or TUDs – in order to make recommendations about ways to make future InVEST recreation model versions account for non-conforming parks. We researched the locations of these parks and produced maps showing the unique combination of location, history, and attributes that may explain the failure of InVEST to predict their visitation.

At the same time as we were analyzing the Tacoma data we valiantly pursued any last ways to infer visitation data at Seattle parks. Eric Asp at Seattle Parks and Recreation suggested the possibility of looking at garbage production or toilet paper use as a way to infer visitation data at parks. Brenda

Gigliotti, Administrative Staff Analyst at Seattle Parks and Recreation, provided us with a table of garbage collection from Seattle city parks. Unfortunately Seattle Parks and Recreation has a centralized transfer station and garbage from each park is not recorded.

3.2 SEATTLE PARKS DATA ANALYSIS

Since the focus of the project was to examine the qualities that attract visitors to Seattle parks, the analysis of park features continues with the Seattle park locations. The city of Seattle has 419 parks ranging in size as shown in **Figure 7**.

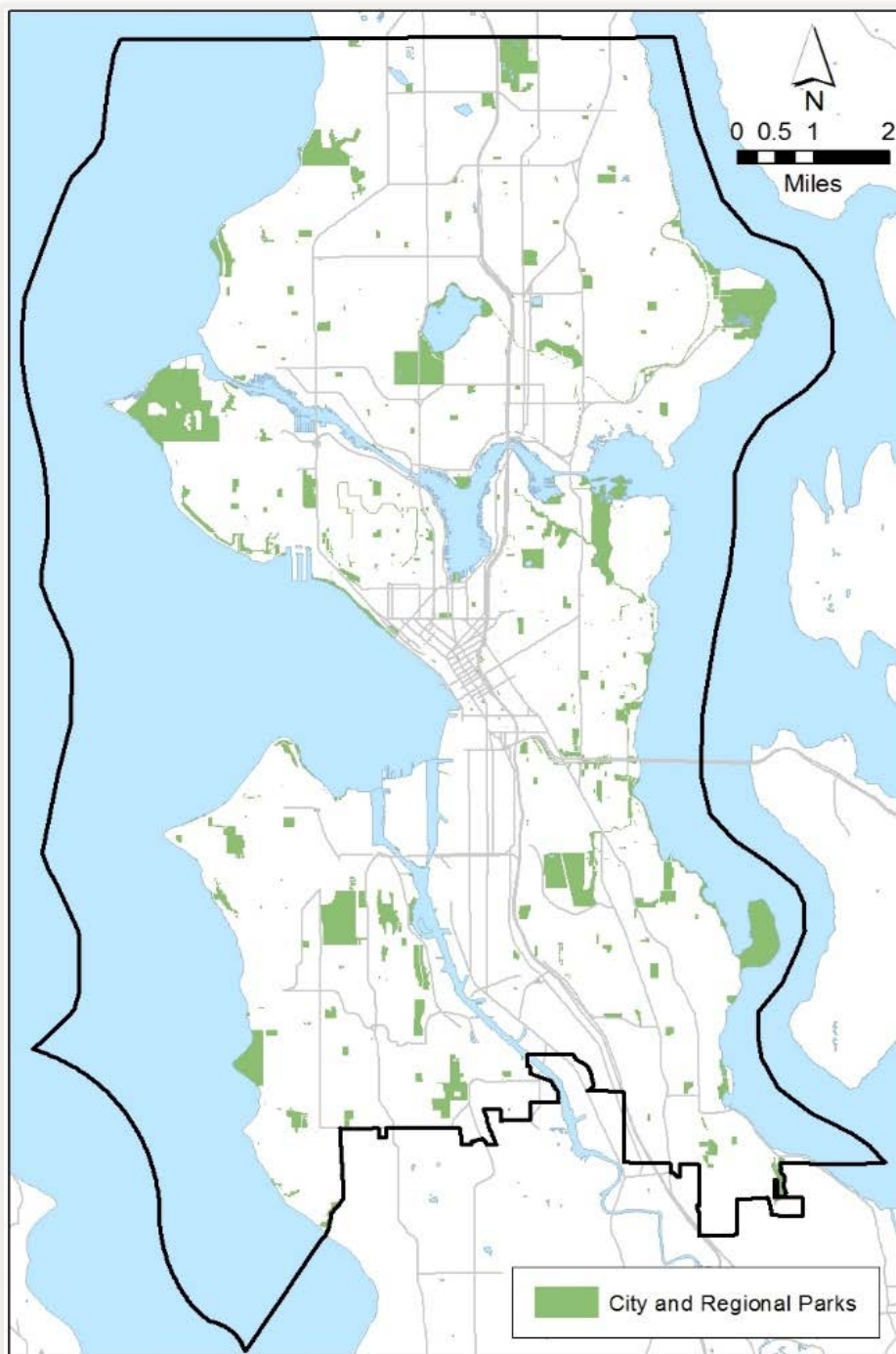


Figure 7. Parks across the City of Seattle.

Data collection was focused on four categories of data: park characteristics and land cover, facilities available at the park, alternative ways of accessing the parks (such as bike paths) and neighborhood demographics. Seven main sources of data were found: King County GIS Data Portal (King County, 2015), City of Seattle (Seattle, 2015), OpenStreetMap (Open Street Map, 2015), NOAA Land Cover data (NOAA, 2010, and Earth Corps (Seattle Urban Nature Project, 2002), as well as the Twitter User Days (TUD) and Flickr Photo User Days (PUD) from the Natural Capital Project as shown **Table 3**. As much data was gathered as possible for predictor variables to be used in any future analysis.

Table 3. Sources of Data Acquired.

Source	Type of Data Acquired
King County	Transportation Data, Census Data, Additional Facilities Data
City of Seattle	Park Facilities
Open-StreetMap	Park Facilities
EarthCorps	Habitat Types
NOAA C-CAP	Land Cover
Natural Capital Project	Twitter User Days (TUD), Flickr Photo User Days (PUD)

Park Characteristics and Land Cover

The data search initially began with the King County GIS Data Portal (King County, 2015). This provided most of the built attributes of the King County parks description and size. Habitat data was initially obtained from EarthCorps as they provided a detailed study across the Seattle area in 2000. We found from EarthCorps’ interactive map (EarthCorps, 2015) that they have a wealth of information on plant species and habitat types in Seattle parks. EarthCorps shared a database containing all their data. Because EarthCorps data did not cover all of the parks it was supplemented with NOAA C-CAP land cover data with a broader range but a more coarse resolution (NOAA, 2012).

Facilities

Seattle and OpenStreetMap (OSM) provided many of the more detailed parks facilities, such as swimming pools, restrooms, play areas, sports facilities, and dog parks. Seattle data was primarily used with OSM supplementing where data was not available. Seattle data was extracted using an available table of facilities with latitude and longitude data and grouped by facility type (Seattle, 2015). A point shapefile was created and a script was used to extract each data type (see **Appendix A**). The OSM data was mostly polygon files that were converted to point files to be used as counts.

Only facilities data can be run through the InVEST Recreation model given the nature of the data. The InVEST model will calculate counts for points, lengths for lines, and areas for polygons intersecting the specified grid area. Since other data incorporated into polygons (such as demographic data and distances to locations) is not appropriate for this, only the facilities data was used in the InVEST model. However, since the Flickr and Twitter data was acquired through the Natural Capital Project, all of the data was analyzed in post-processing instead of through the InVEST recreation model.

Accessibility to Alternative Transportation

King County had information on county-wide alternative transportation systems, including bus stops, trails, and bike paths. Distance to bus stops, trails, and bike paths were calculated and the number of bus stops within a quarter mile of the park were also calculated as possible predictor variables.

Neighborhood Demographics

King County also had consolidated 2010 census data for the county which was used to determine neighborhood demographics. Demographic data was broken out into population density, median household income, percent white, percent black, percent Asian, percent Hispanic, and percent Native American.

The proposed variables and some descriptive statistics of each are shown in **Table 4**. None of the variables are normally distributed.

Table 4. Proposed predictor variables grouped by category

Predictor Variable Group	Source	Predictor Variable Name	Predictor Variable	Mean	St Dev	Min	Max
Park Characteristics and Land Cover	King County	Shape Area (m ²)	Shape_Area	51,776	173,139	17	2,132,024
	King County	(presence/absence in 0.2 mi buffer)	WaterPres	0.4	0.5	0	1
	Seattle	Waterfront (count)	Waterfront	0.2	0.4	0	3
	Earth Corps	Percent Water (%)	WaterPer	1	6	0	78
	Earth Corps	Percent Forest (%)	ForestPer	39	36	0	100
	Earth Corps	Percent Grass (%)	GrassPer	16	26	0	100
	Earth Corps	Percent Developed (%)	DevPer	21	27	0	100
	NOAA	Percent Water (%)	NOAAWaterPer	4	13	0	84
	NOAA	Percent Forest (%)	NOAAForestPer	16	26	0	100
	NOAA	Percent Grass (%)	NOAAGrassPer	9	15	0	75
	NOAA	Percent Developed (%)	NOAADevPer	66	34	0	100
Facilities and Features	OSM	BBQ (count)	BBQ	0.04	0.40	0	5
	Seattle	Boat Launch (count)	BoatLaunch	0.09	0.39	0	4
	Seattle	Children's Play Area (count)	PlayArea	0.28	0.49	0	3
	Seattle	Community Center (count)	CommCenter	0.05	0.21	0	1
	Seattle and OSM	Count of Beach (count in 1/4 mi buffer)	Beach	0.04	0.20	0	2
	Seattle	Dog Park (count)	OffLeash	0.03	0.16	0	1
	OSM	Fields (count)	Field	0.28	0.97	0	9
	Seattle	Fishing (count)	Fishing	0.02	0.14	0	1
	Seattle and KC	Garden (count)	Gardens	0.02	0.15	0	2
	OSM	Hard Courts (count)	HardCourt	0.28	0.75	0	6
	OSM	Parking Area (count)	Parking	0.41	1.33	0	13
	Seattle	Picnic Areas (count)	Picnic	0.06	0.24	0	1
	OSM	Sports Facilities (count)	SportFac	0.69	1.91	0	18
	Seattle	Swimming Pool (count)	Pool	0.01	0.10	0	1
	OSM	Toilets (count)	Toilets	0.12	0.62	0	10
	Seattle	View (count)	View	0.23	0.42	0	1
Seattle	Wading Pool (count)	WadingPool	0.07	0.25	0	1	
Accessibility to Alternative Transportation	King County	Distance to Trail (100m)	Dist2Trail	8.28	8.90	0	51
	King County	Distance to Bike Lanes (100m)	Dist2Bike	0.55	0.90	0	5
	King County	Distance to Bus Stop (100m)	Dist2Bus	1.05	1.26	0	8
	King County	Bus Stop (count in 1/4 mile buffer)	BusStop	16	15	0	178
Neighborhood Demographics (within a Quarter Mile of Park)	2010 Census/ King County	Population Density (people/1000m ²)	PopDensity	5	4	0	24
	2010 Census/ King County	Percent Asian and Pacific Islander (%)	PerAsian	12	10	0	60
	2010 Census/ King County	Percent Black (%)	PerBlack	8	9	0	36
	2010 Census/ King County	Percent Hispanic (%)	PerHispanic	6	5	0	40
	2010 Census/ King County	Percent American Indian/Alaska Native (%)	PerNative	1	2	0	18
	2010 Census/ King County	Percent White (%)	PerWhite	72	19	15	96
	2010 Census/ King County	Percent Not White (%)	PerNotWhite	28	19	4	85
	2010 Census/ King County	Median Household Income (USD)	HHIncome	68,627	23,216	20,275	138,472

Twitter Data and Flickr Data

The Twitter and Flickr datasets were obtained directly from the Natural Capital Project which was able to extract averages for both Tacoma and King County along with the data used in the Verification Analysis. For the Seattle comparison, the averages across the entire dataset of either Flickr or Twitter were used for the comparison to the predictor variables.

Data Processing

In the interest of making the project reproducible and extendable, scripts were written for many of the geoprocessing operations. Initially scripts were produced to clip the data down to Seattle but the Natural Capital Project wished to keep all data within the final product for further evaluation.

Land Use and Habitat data was aggregated in fewer categories as defined in **Appendix B**. Four grouped categories of interest were defined: Water, Forest, Grass/Open Space, and Developed. The area was intersected with the parks layer and a percent of the total park area was calculated for each location.

The park features and facilities were added to the parks shapefile using a script which iterated over a spatial join found in **Appendix A-3**. This summed the number of points for each feature type in each park and created a new column with this information.

For analysis of the transportation features near each park, the Near tool was used to determine the distance to each feature. For the number of bus stops, a quarter mile buffer was created and spatial join was used to sum the number of points in each feature.

For determining the demographic data a quarter mile buffer was used as the neighborhood area near the park. Larger sized buffers were also produced but were not finished for analysis for this report. A quarter mile walking distance is typical of most of the smaller parks and many of the park neighborhoods overlap even at that distance. Further exploration would examine larger neighborhood regions for larger parks.

The quarter mile buffer was intersected with the demographic data and a geometric average was calculated based on the area of each census block that was within each park buffer area. The percentage of each category was based on this calculated value.

The final results were all produced in tables for each data type and linked to original parks data source by a unique park ID to create a matrix that could be used for analysis.

A correlation matrix was used to examine correlations between each group of variables. The feature class variables were not continuous values and so a non-parametric t-test was used to compare the data.

4 RESULTS

The follow results were produced for the Tacoma Parks validation and the analysis of the Seattle Parks feature attributes.

4.1 TACOMA PARKS VALIDATION

Scatter plots of park visitation against mean Flickr Photo User Days (PUD) and Twitter User Days (TUD) for 2012 to 2014 are presented in **Figure 8** below.

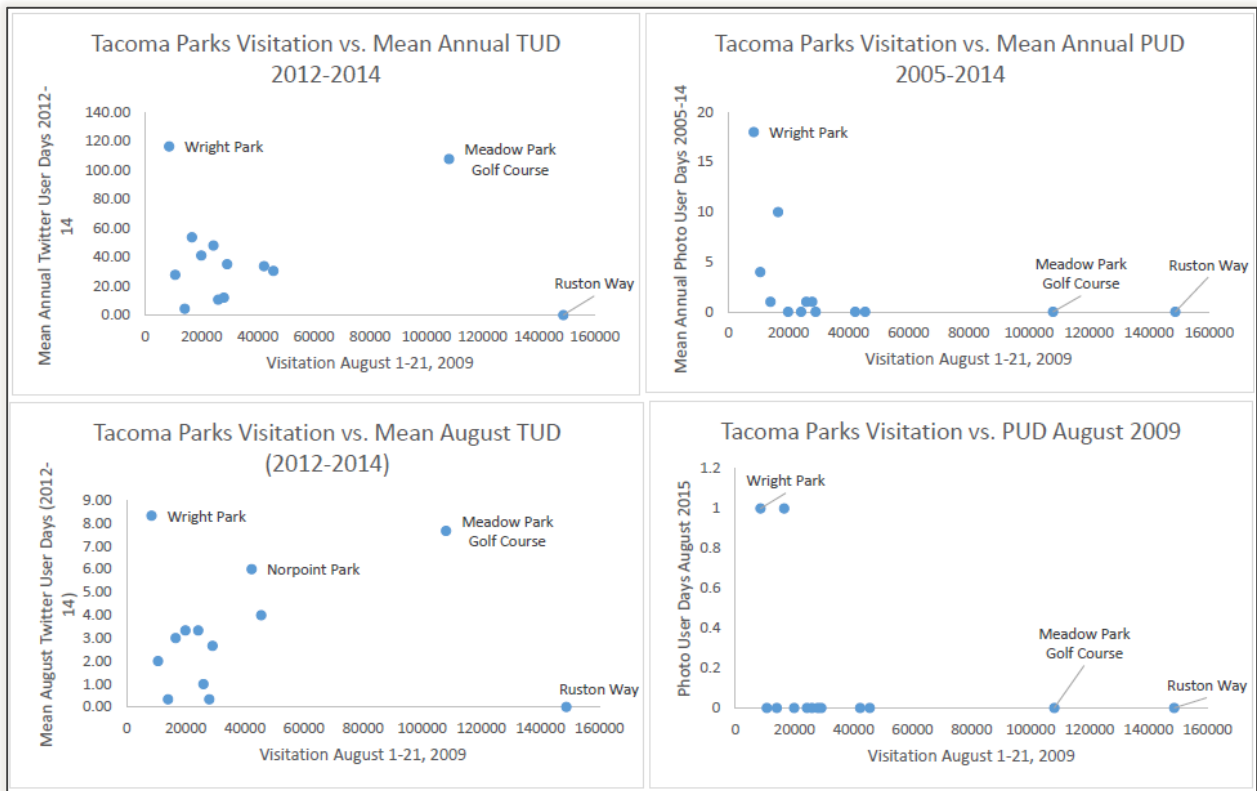


Figure 8. Scatter plots of park visitation against PUD and TUD (Point Defiance Park omitted)

Since most of the photo user days are zero, the primary dataset for examination will be the TUD values. A comparison of the Mean TUD for August with a trendline is shown in **Figure 9a** along with Ruston Way removed from the dataset in **Figure 9b**.

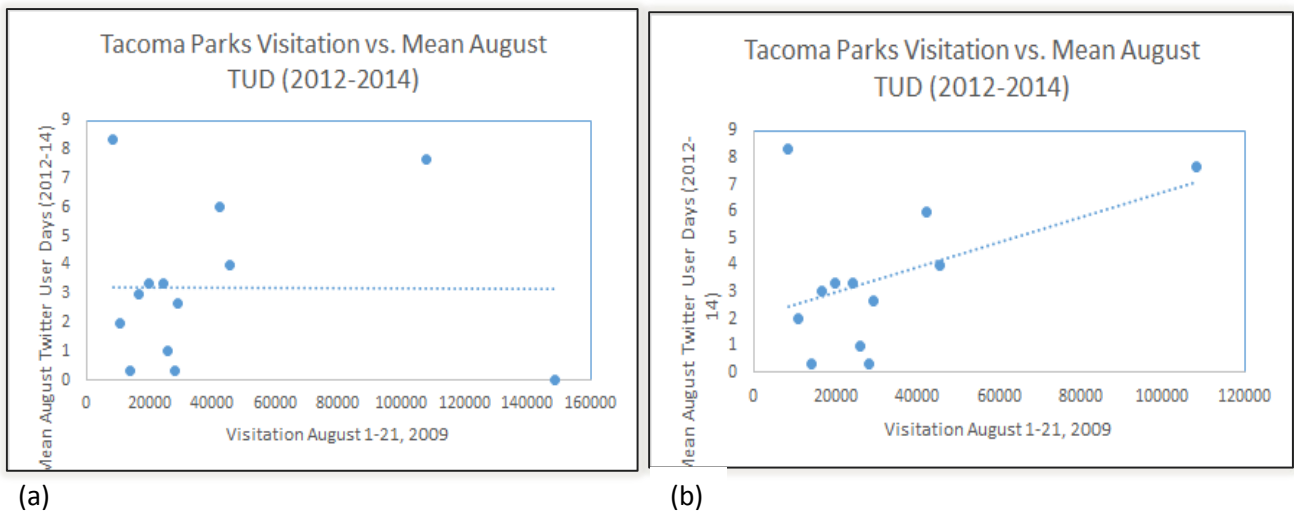


Figure 9. (a) Trendlines for August TUD against visitation and (b) the same data with the outlier removed.

4.2 SEATTLE PARKS

A table was created to compare all of the proposed variables for each of the parks and was limited to the Seattle dataset. Correlation matrices were used to examine correlations between different sets of variables to determine which variables were the most promising. None of the variables are normally distributed as can be seen in **Table 4**, and there are many zero values, particularly for the smaller parks.

The first comparison was between the Twitter User Days (TUD) mean and the Photo User Days (PUD) along with the shape area to see what comparisons could be made. As shown in **Figure 10**, the TUD and PUD are statistically significant and a Pearson Correlation coefficient of 0.88. The shape area is also significant but does not have as high of a Pearson Correlation coefficient at 0.53.

Correlations

		TUD_mean	mean_PUD	Shape_Area
TUD_mean	Pearson Correlation	1	.882**	.527**
	Sig. (2-tailed)		.000	.000
	N	419	419	419
mean_PUD	Pearson Correlation	.882**	1	.513**
	Sig. (2-tailed)	.000		.000
	N	419	419	419
Shape_Area	Pearson Correlation	.527**	.513**	1
	Sig. (2-tailed)	.000	.000	
	N	419	419	419

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 10. Results of comparison of TUD, PUD, and Shape Area data.

Figure 11 shows a comparison between TUD and PUD to the Habitat and Land Cover variables identified previously. While there are several variables that are statistically significant, most have a low r-squared value. The largest Pearson coefficient is between TUD and the percent water calculated from the EarthCorps dataset.

Correlations

		TUD_mean	mean_PUD	tbl_EarthCorpsHabitat_PerWater	tbl_EarthCorpsHabitat_PerForest	tbl_EarthCorpsHabitat_PerGrass	tbl_EarthCorpsHabitat_PerDev	tbl_NOAAHabitatPercent_PerWater	tbl_NOAAHabitatPercent_PerGrass	tbl_NOAAHabitatPercent_PerForest	tbl_NOAAHabitatPercent_PerDev
TUD_mean	Pearson Correlation	1	.882**	.326**	-.106	-.038	.079	.168**	.092	-.022	-.108*
	Sig. (2-tailed)		.000	.000	.075	.525	.185	.001	.060	.649	.027
	N	419	419	282	282	282	281	418	418	418	417
mean_PUD	Pearson Correlation	.882**	1	.132*	-.070	-.051	.082	.127**	.092	.028	-.154**
	Sig. (2-tailed)	.000		.027	.241	.395	.172	.009	.059	.566	.002
	N	419	419	282	282	282	281	418	418	418	417
tbl_EarthCorpsHabitat_PerWater	Pearson Correlation	.326**	.132*	1	-.079	-.057	.018	.246**	-.047	-.053	-.016
	Sig. (2-tailed)	.000	.027		.188	.340	.761	.000	.434	.375	.790
	N	282	282	282	282	282	281	282	282	282	281
tbl_EarthCorpsHabitat_PerForest	Pearson Correlation	-.106	-.070	-.079	1	-.456**	-.531**	-.186**	-.087	.498**	-.323**
	Sig. (2-tailed)	.075	.241	.188		.000	.000	.002	.145	.000	.000
	N	282	282	282	282	282	281	282	282	282	281
tbl_EarthCorpsHabitat_PerGrass	Pearson Correlation	-.038	-.051	-.057	-.456**	1	-.041	-.151*	.381**	-.320**	.235**
	Sig. (2-tailed)	.525	.395	.340	.000		.492	.011	.000	.000	.000
	N	282	282	282	282	282	281	282	282	282	281
tbl_EarthCorpsHabitat_PerDev	Pearson Correlation	.079	.082	.018	-.531**	-.041	1	-.107	-.026	-.300**	.342**
	Sig. (2-tailed)	.185	.172	.761	.000	.492		.073	.659	.000	.000
	N	281	281	281	281	281	281	281	281	281	281
tbl_NOAAHabitatPercent_PerWater	Pearson Correlation	.168**	.127**	.246**	-.186**	-.151*	-.107	1	-.047	-.142**	-.331**
	Sig. (2-tailed)	.001	.009	.000	.002	.011	.073		.340	.004	.000
	N	418	418	282	282	282	281	418	418	418	417
tbl_NOAAHabitatPercent_PerGrass	Pearson Correlation	.092	.092	-.047	-.087	.381**	-.026	-.047	1	-.086	-.348**
	Sig. (2-tailed)	.060	.059	.434	.145	.000	.659	.340		.077	.000
	N	418	418	282	282	282	281	418	418	418	417
tbl_NOAAHabitatPercent_PerForest	Pearson Correlation	-.022	.028	-.053	.498**	-.320**	-.300**	-.142**	-.086	1	-.696**
	Sig. (2-tailed)	.649	.566	.375	.000	.000	.000	.004	.077		.000
	N	418	418	282	282	282	281	418	418	418	417
tbl_NOAAHabitatPercent_PerDev	Pearson Correlation	-.108*	-.154**	-.016	-.323**	.235**	.342**	-.331**	-.348**	-.696**	1
	Sig. (2-tailed)	.027	.002	.790	.000	.000	.000	.000	.000	.000	
	N	417	417	281	281	281	281	417	417	417	417

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Figure 11. Results of comparisons between TUD, PUD, and Habitat/Land Cover data.

		TUD_mean	mean_PUD	PopDensity (per 1000 m2)	PerWhite	PerBlack	PerNative	PerAsian	PerHispanic	PerNotWhite	MHHI
TUD_mean	Pearson Correlation	1	.882**	.173**	.092	-.077	-.013	-.073	-.081	-.092	-.031
	Sig. (2-tailed)		.000	.000	.059	.118	.785	.136	.099	.059	.521
	N	419	419	419	419	419	419	419	419	419	419
mean_PUD	Pearson Correlation	.882**	1	.129**	.127**	-.112*	-.025	-.091	-.103*	-.127**	.017
	Sig. (2-tailed)	.000		.008	.009	.022	.612	.061	.035	.009	.733
	N	419	419	419	419	419	419	419	419	419	419
PopDensity (per 1000 m2)	Pearson Correlation	.173**	.129**	1	-.076	.043	.181**	.087	-.023	.076	-.493**
	Sig. (2-tailed)	.000	.008		.118	.375	.000	.076	.635	.118	.000
	N	419	419	419	419	419	419	419	419	419	419
PerWhite	Pearson Correlation	.092	.127**	-.076	1	-.809**	-.311**	-.869**	-.436**	-1.000**	.518**
	Sig. (2-tailed)	.059	.009	.118		.000	.000	.000	.000	.000	.000
	N	419	419	419	419	419	419	419	419	419	419
PerBlack	Pearson Correlation	-.077	-.112*	.043	-.809**	1	.151**	.475**	.203**	.809**	-.331**
	Sig. (2-tailed)	.118	.022	.375	.000		.002	.000	.000	.000	.000
	N	419	419	419	419	419	419	419	419	419	419
PerNative	Pearson Correlation	-.013	-.025	.181**	-.311**	.151**	1	.182**	.248**	.311**	-.373**
	Sig. (2-tailed)	.785	.612	.000	.000	.002		.000	.000	.000	.000
	N	419	419	419	419	419	419	419	419	419	419
PerAsian	Pearson Correlation	-.073	-.091	.087	-.869**	.475**	.182**	1	.311**	.869**	-.458**
	Sig. (2-tailed)	.136	.061	.076	.000	.000	.000		.000	.000	.000
	N	419	419	419	419	419	419	419	419	419	419
PerHispanic	Pearson Correlation	-.081	-.103*	-.023	-.436**	.203**	.248**	.311**	1	.436**	-.397**
	Sig. (2-tailed)	.099	.035	.635	.000	.000	.000	.000		.000	.000
	N	419	419	419	419	419	419	419	419	419	419
PerNotWhite	Pearson Correlation	-.092	-.127**	.076	-1.000**	.809**	.311**	.869**	.436**	1	-.518**
	Sig. (2-tailed)	.059	.009	.118	.000	.000	.000	.000	.000		.000
	N	419	419	419	419	419	419	419	419	419	419
MHHI	Pearson Correlation	-.031	.017	-.493**	.518**	-.331**	-.373**	-.458**	-.397**	-.518**	1
	Sig. (2-tailed)	.521	.733	.000	.000	.000	.000	.000	.000	.000	
	N	419	419	419	419	419	419	419	419	419	419

Figure 12. Results of comparison between TUD, PUD, and demographic variables.

A similar comparison was done using the demographic data as shown in **Figure 12**. The most statistically significant variable is the population density but the Pearson Correlation Coefficient is low at only around 0.17 and 0.12 respectively for TUD and PUD dependent variables. These variables also show a positive correlation with higher percent white populations but a negative one with other demographics, although slight.

A correlation matrix was set up between the accessibility variables and the TUD and PUD outputs as shown in **Figure 13**. All of the variables are statistically significant, but the Pearson Coefficient is small and negative for distances to transportation options and is positive for the number of bus stops (BusStop).

Correlations

		TUD_mean	mean_PUD	Shape_Area	Dist2Trail	Dist2Bike	Dist2Bus	BusStop
TUD_mean	Pearson Correlation	1	.882**	.527**	-.178**	-.156**	-.156**	.350**
	Sig. (2-tailed)		.000	.000	.000	.001	.001	.000
	N	419	419	419	419	419	419	419
mean_PUD	Pearson Correlation	.882**	1	.513**	-.172**	-.145**	-.092	.171**
	Sig. (2-tailed)	.000		.000	.000	.003	.061	.000
	N	419	419	419	419	419	419	419
Shape_Area	Pearson Correlation	.527**	.513**	1	-.088	-.128**	-.159**	.211**
	Sig. (2-tailed)	.000	.000		.071	.008	.001	.000
	N	419	419	419	419	419	419	419
Dist2Trail	Pearson Correlation	-.178**	-.172**	-.088	1	.253**	.064	-.211**
	Sig. (2-tailed)	.000	.000	.071		.000	.188	.000
	N	419	419	419	419	419	419	419
Dist2Bike	Pearson Correlation	-.156**	-.145**	-.128**	.253**	1	.229**	-.170**
	Sig. (2-tailed)	.001	.003	.008	.000		.000	.000
	N	419	419	419	419	419	419	419
Dist2Bus	Pearson Correlation	-.156**	-.092	-.159**	.064	.229**	1	-.404**
	Sig. (2-tailed)	.001	.061	.001	.188	.000		.000
	N	419	419	419	419	419	419	419
BusStop	Pearson Correlation	.350**	.171**	.211**	-.211**	-.170**	-.404**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	419	419	419	419	419	419	419

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 13. Results of comparison between TUD, PUD, and Accessibility variables.

Table 5 below shows the analysis of the parks feature class variables using a Mann Whitney U t-test to compare means of the PUDs with the attribute present and without. This was done since there is not enough variation for it to be considered continuous; usually a park feature is either present (1) or it is not (0). The table is ordered by the mean PUD for each attribute. All features with significant differences are shown to have a mean greater when the attribute is present. This is indicated by the greater than (>) symbol in the last column.

Similarly, **Table 6** shows the test performed using the Twitter dataset for comparison.

Table 5. Comparison of samples means of Flickr PUD where an attribute is present or not using a Mann Whitney U t-test.

Attributes Present at Parks	Number of Parks where Present	Mean With Attribute	Mean Without Attribute	Average Picture User Days (Flickr)
Beach	14	63	7	>
Fishing	9	61	8	>
OffLeash	11	52	8	>
Picnic	26	50	6	>
Toilets	33	48	5	>
Pool	4	43	8	>
Boat Launch	29	37	7	>
Parking	78	28	4	>
Waterfront	56	26	6	>
View	97	21	5	>
Wading Pool	28	18	8	>
Field	52	18	8	>
Play Area	110	18	6	>
Total Sports Facilities	96	18	6	>
Water Present	179	16	3	>
HardCourt	69	16	8	>
Community Center	20	10	9	>
Gardens	7	80	8	
BBQ	9	42	8	

Note: > indicated mean with attribute is statistically significant

Table 6. Comparison of samples means of Twitter User Days (TUD) where an attribute is present or not using a Mann Whitney U t-test

Attributes Present at Parks	Number of Parks where Present	Mean With Attribute	Mean Without Attribute	Average Twitter User Days
Pool	4	372	43	>
Beach	14	317	36	>
Fishing	9	306	40	>
Toilets	33	222	31	>
Picnic	26	215	35	>
Boat Launch	29	179	36	>
Parking	78	129	27	>
Waterfront	56	119	34	>
Wading Pool	28	116	41	>
Community Center	20	104	43	>
Field	52	100	38	>
Total Sports Facilities	96	99	30	>
Hard Court	69	92	37	>
Play Area	110	86	31	>
View	97	83	34	>
Water Present	179	74	25	>
Gardens	7	261	42	
Off Leash	11	259	40	
BBQ	9	224	43	

Note: > indicated mean with attribute is statistically significant

The final distribution of PUDs across Seattle parks are shown in **Figure 14** below. This data is not normally distributed with most of the parks showing few or no user days and a few parks have a larger number of user days. Similarly, TUD shown in **Figure 15** is skewed with most parks showing less than 10 annual average user days and a few showing up to 1100 average annual photo user days.

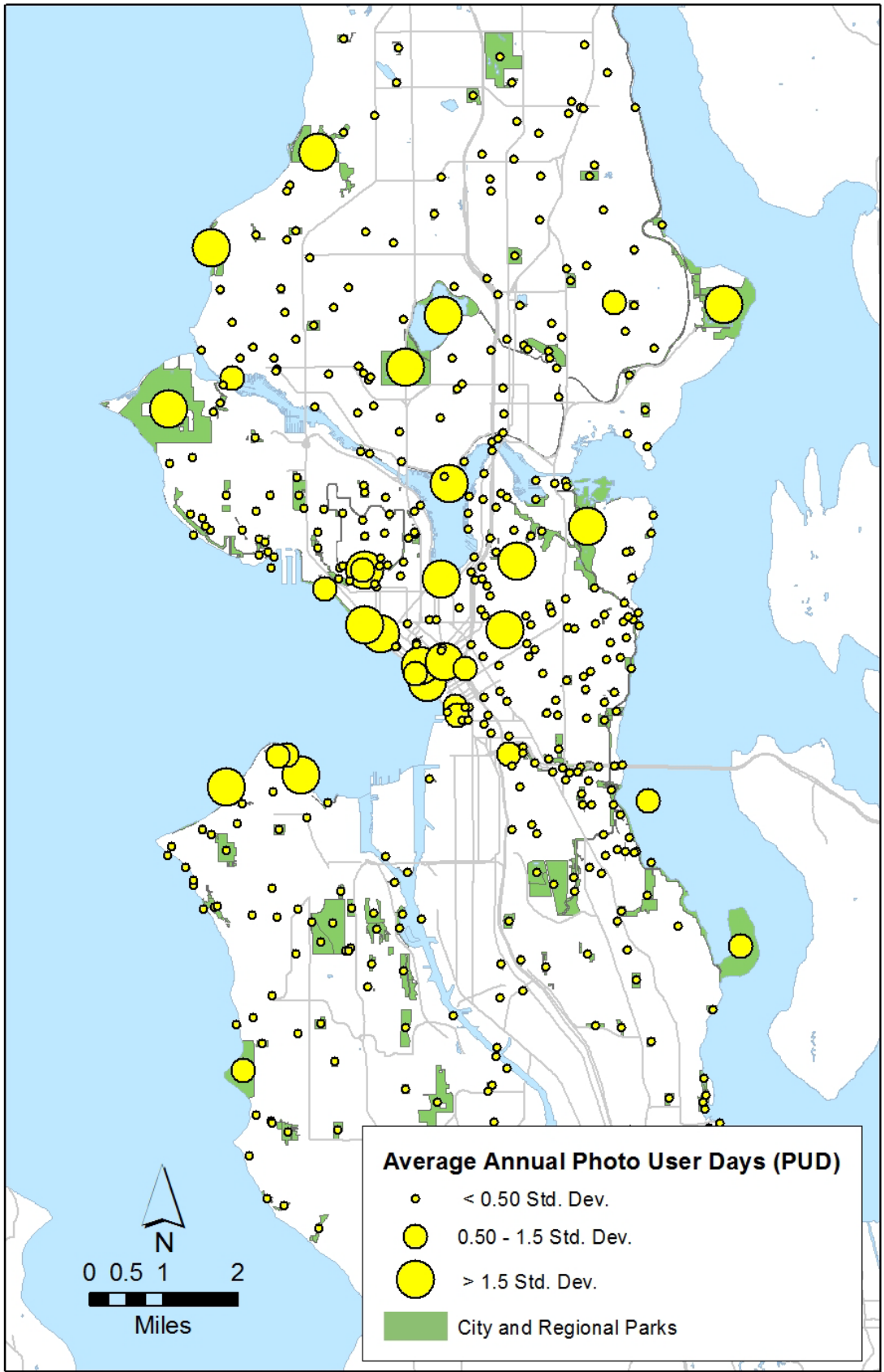


Figure 14. Distribution of Average Annual Photo User Days (Flickr) in Seattle Parks

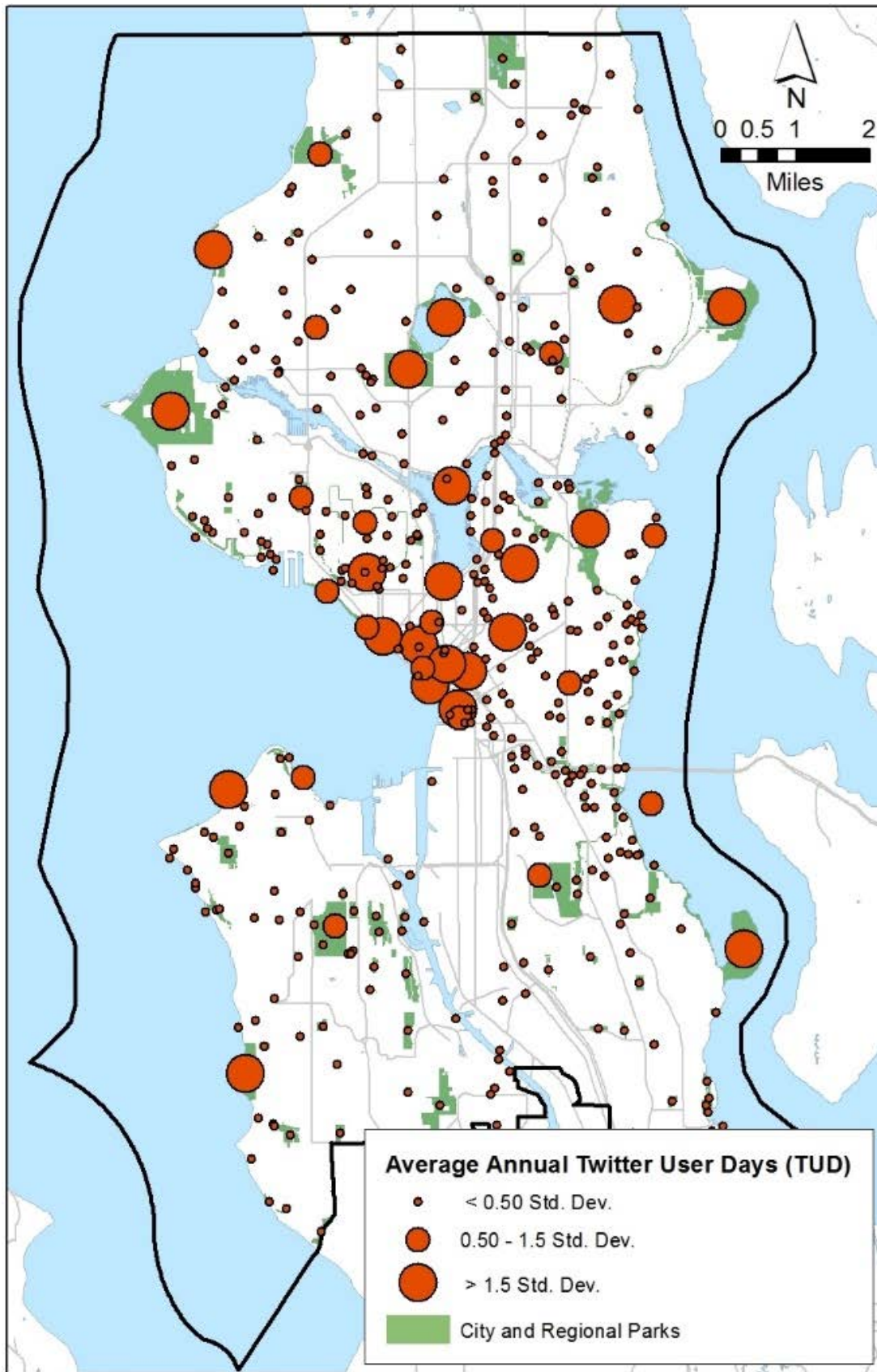


Figure 15. Distribution of Average Annual Twitter User Days in Seattle Parks

5 DISCUSSION

The following discussion covers both the Tacoma Validation analysis and the Seattle Parks data analysis.

5.1 TACOMA VALIDATION DISCUSSION

When viewing the scatter plots (**Figure 8**) it becomes immediately apparent that there are not enough Flickr photo user days to show a meaningful relationship between PUD and visitation. Thus we are unable to comment on how well PUDs predict visitation in Tacoma except to note that Wright Park Arboretum, in addition to having a high proportion of TUDs also has the highest rate of PUDs.

On the other hand the TUD values are more complete and the annual and August TUD scatter plots show similar trends, indicating that the August mean is representative of the yearly pattern and therefore useful in comparing against visitation data. Ruston Way (bottom right of the plot), which has the most visitors but no TUDs, appears as an outlier. Ruston Way is a linear park, popular with walkers and joggers (Ruston Way Park) and therefore may not have a lot of visitors spending enough time not moving in order to Tweet.

The R Square and trend lines in **Figure 9** tell us that TUDs are not a valid predictor of visitation when Ruston Way is included in the calculation (**Figure 9a**). Removing Ruston Way (**Figure 9b**) from the data results in a positive correlation between TUDs and visitation, but still with a small R Square value, indicating that Twitter data explains some of the variation in visitation.

At this point we decided it would be beneficial to examine the types of parks for which the TUD/PUD visitation proxy works well and types for which it works poorly in order to give recommendations about future InVEST model development.

We discovered that the model grossly overestimated visitation for Wright Park Arboretum and Conservatory (**Figure 14**). This is a park with a high number of both PUDs and TUDs, but low actual visitation. The park contains a historic 1907 conservatory, and the arboretum was designed by a famous Bavarian architect of the time. It is in the center of a commercial neighborhood close to the city center and port. Possible explanations for the inaccuracy of PUD/TUD proxy data in such a park are that historical parks attract a limited audience but one that takes the time to post photos and to tweet, this could be a picturesque park in a less desirable neighborhood and people go there for weddings and events but not for recreation, or the problem could be caused by incorrectly geotagged tweets and photos from the surrounding commercial district.

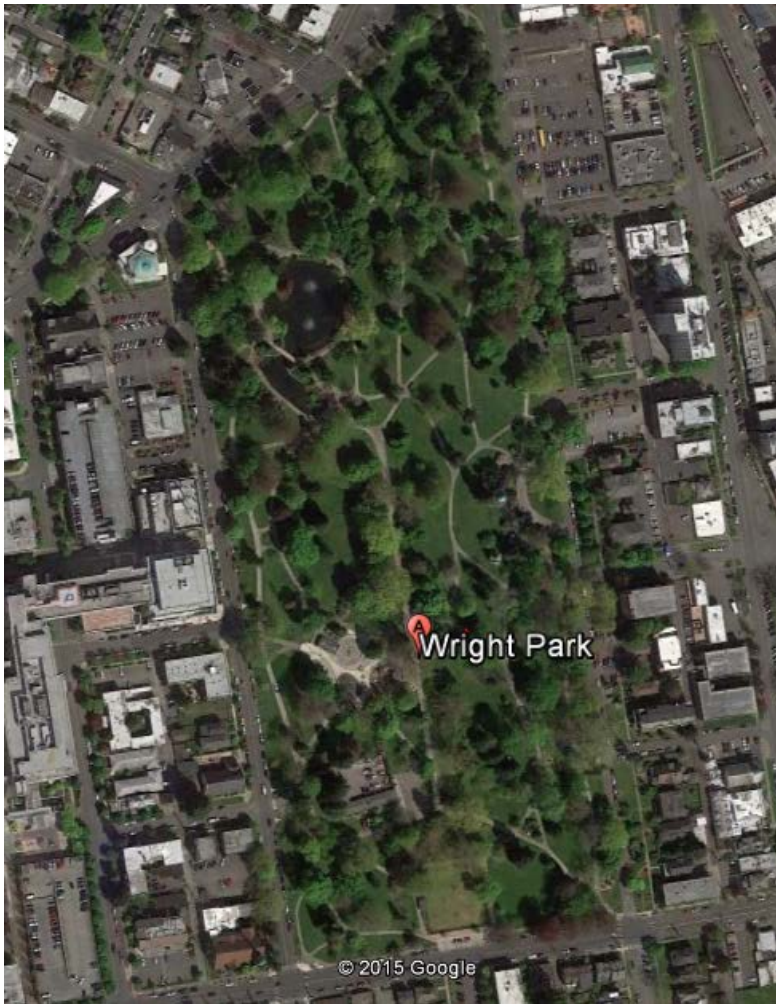


Figure 16. Wright Park Arboretum and Conservatory

The model worked well for Meadow Park Golf Course in **Figure 15** as it has both high visitation and high TUDs (there is not enough PUD data to comment upon). This indicates that the rate of Tweeting at public golf courses is on par with the general rate of tweeting at urban parks.

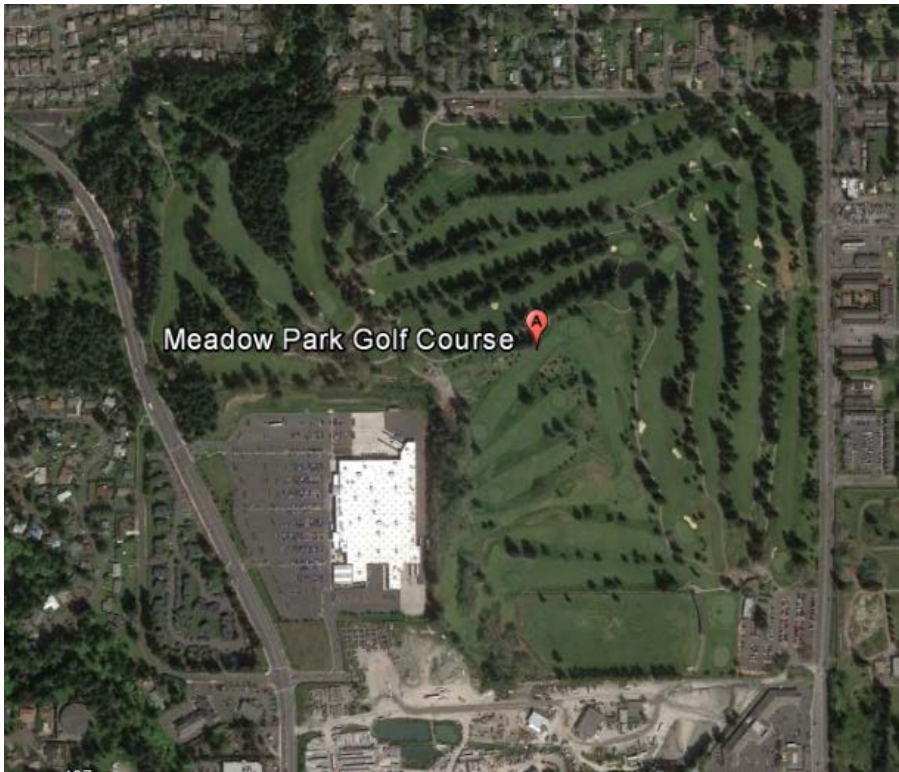


Figure 17. Meadow Park Golf Course

Another outlier for which the model did not work well for is Ruston Way Park (**Figure 16**), one with very high visitation but low TUDs.

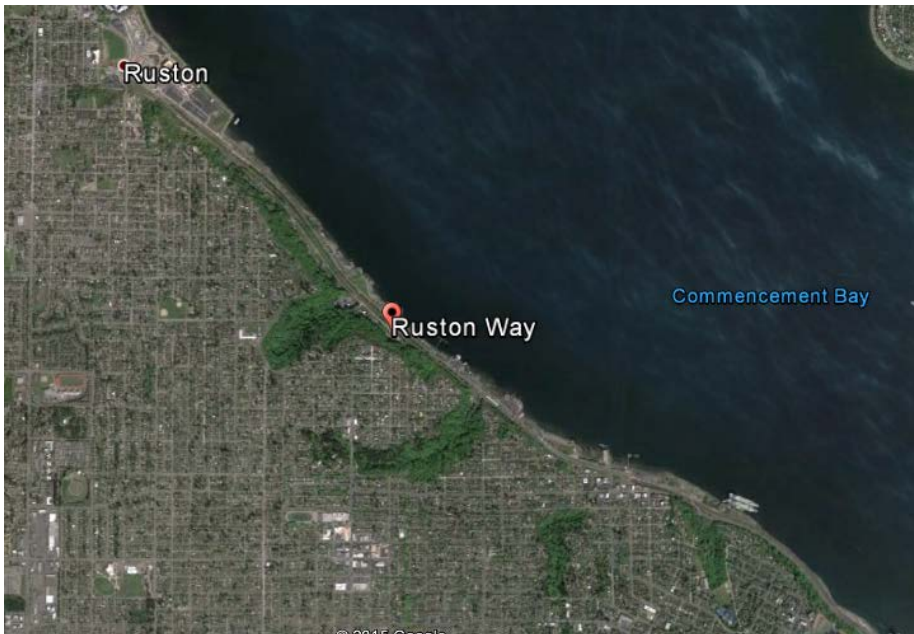


Figure 18. Ruston Way Escarpment. Ruston Way Park is one of a series of parks along the walking and biking trail that follows the coastline.

Ruston Way is one of many parks along a paved waterfront recreation trail. The number of TUDs and PUDs may be low because users are walking along the path but not stopping long enough to Tweet or

take photos as they have a choice of many parks in which to stop along their path. This type of park is likely to have its visitation underestimated by InVEST or other proxy based models and it shows that when using such a model it is wise to research park type and make adjustments.

5.2 SEATTLE PARK FEATURE ANALYSIS

Like the Tacoma parks analysis, many of the parks have little or no visitation proxy data presented by PUD and TUD values, making analysis difficult. Likewise, none of the chosen variables are normally distributed and are all heavily skewed. However, the strong Pearson Correlation coefficient of 0.88 between the TUD and PUD values across all parks suggest that they are showing similar visitation patterns even though slightly different demographics are being represented by each dataset. The area of each park also seems to play a contributing factor which would be expected. This also might suggest that further analysis should include developing separate datasets by park size and usage.

The correlation coefficients for park habitat and land cover data do not show strong correlations between types as well as for the TUD and PUD variables. This is unexpected since a correlation between landcover results for the EarthCorp data versus the NOAA data was expected but not demonstrated, the strongest Pearson Coefficient being 0.5 between the two percent forest datasets. While there should be some variation, this is unfortunate since it reflects that the NOAA data is not as meaningful as the more detailed EarthCorp dataset. Most likely this is skewed by the smaller parks being listed as entirely developed under the coarse resolution of the NOAA data. Further detailed habitat/land cover data might be necessary if these variables are going to be investigated further, especially for smaller urban parks.

While most of the demographic data did not show a significant correlation as shown in **Figure 12**, the population density and percent white were significant for both variables although they would appear to not have a significant impact on visitation. The analysis also shows a trend of percent white being a positive correlation while the other demographic groups represent a negative correlation. This would be a relationship worth investigating as how different groups use parks or if parks in whiter neighborhoods are considered more “valuable”.

The correlation matrix in **Figure 13** does show a statistically significant correlation between all variables for both TUD and PUD, although small. As expected, the distance calculations are showing a negative correlation (as distances to transportation grow, the visitation drops) while the number of bus stops in the vicinity shows a positive correlation. Further investigation in the distances that different groups of people are willing travel should be investigated.

The non-parametric t-test shown in **Table 5** and **Table 6** suggests that different park attributes do attract visitors. On the top of the list for Flickr PUD data are beaches, dog parks, picnic areas, and restroom facilities. Fishing is also at the top but only a few parks have fishing areas and these are most likely reflected in the beach areas, also. The top of the TUD list are beaches, restrooms, picnic areas, and boat launch areas. Pools is at the very top but there are not very many locations listed with pools and it might not be as meaningful. Dog parks has fallen off the list and may reflect the difference between Flickr (known to be slightly older on average) and Twitter users. Beaches and Restroom facilities are also top of both list indicating that these are features that attract more visitors. The evidence of the popularity of beaches is shown in Figure 14 and Figure 15 where the higher Twitter and flicker results are prominent. Restroom facilities probably correlate heavily to the size of the park also. These results

suggest that some park facilities and features do have an impact on visitation and further investigation should be done.

While the correlation results do not show strong indicators of specific variables being more likely to draw visitors than others, it does suggest that some variables, and more likely a combination of variables, do contribute to park visitation rates. Parks size is evidently one of them and further analysis is recommend by breaking parks up by size and usage. Park attributes such as beaches and restroom facilities also appear to have a greater impact on visitation. Further analysis is recommended for all variables in future investigations.

6 BUSINESS CASE AND FURTHER RECOMMENDATIONS

The InVEST toolset is an evolving and effective way of valuing ecosystem services, helping the Natural Capital project fulfil its aim to “integrate the values of nature into all major decisions affecting the environment and human well-being” and to “improve the state of biodiversity and human well-being by motivating greater and more cost-effective investments in both” (NatCap 2015). As such, we recommend the following functionality be added to the InVEST recreation model in order to make it useful to a broader audience:

First, we recommend that the 40,592 square meter (approximately 10 acre) minimum polygon size limitation in the model be eliminated. It could be replaced with a warning message saying that small polygons require more processing time. There are many potential applications that include polygons smaller than this. In fact this study would not have been possible without the Flickr data provided exclusively to us by the Natural Capital Project. Eliminating the minimum polygon size limitation will give the InVEST recreation model a greater appeal to a broader audience and therefore strengthen its role in influencing policy decisions. This project could not have been effectively done without the Flickr data NatCap provided us for the smaller parks.

Secondly, we recommend that the model include the option to calculate visitation based on Twitter data. This project showed that Twitter data is far more plentiful in most location than Flickr data, thus allowing smaller-scale projects to still be able to generate statistically meaningful results. Additionally, Twitter draws from a slightly different demographic than Flickr. Flickr tends to be a slightly more female and older demographic, while Twitter has a younger average age of users and tends to be a little more male-heavy (Fisher, 2015). By receiving both PUDs and TUDs and comparing and/or averaging them, users could get more balanced results.

The results of this project are similar to results in New York City and Minneapolis/St. Paul and we recommend continuing this study on more cities to make the findings more robust. Having similar patterns across US cities supports predictions about the ways that Twitter and Flickr data predict visitation and therefore strengthens support for their use as a general proxy for parks visitation.

We recommend NatCap commission further research into the way that park use, history and location on a trail may skew Flickr and Twitter data, with an eye on building future InVEST models that can recognize and adjust for parks that are likely to be outliers in the TUD/FUD models. This could be done by recognizing park shape and size as a factor or by including a national parks attribute database in the model.

We also recommend that the Natural Capital Project actively research other metrics of parks as an ecosystem service, such as tourism, employment, air quality and resulting health benefits and consider including data from studies that have already been carried out on these subjects. This would allow the model to truly put a value on parks as an ecosystem service and to make a strong economic case that would resonate with decision makers.

7 LITERATURE CITED

"About The Natural Capital Project." Natural Capital Project. Accessed August 23, 2015.

Hebert, Jim, Raechelle Turner, and Hussein Hakam. "Executive Summary Economic Impact Analysis." Phase1Study_Hebert. 2010. Metro Parks Tacoma. Accessed August 15, 2015.

"Ruston Way Park." Wikipedia. Accessed August 16, 2015.

Keeler, B.L., et al. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Front Ecol Environ* 2015; 13(2): 76-81, doi:10.1890/140124 (Jan 2015).

"King County GIS Data Portal." KCGIS Data Download. Accessed August 18, 2015.

"Interactive Habitat Map." EarthCorps. Accessed August 18, 2015.

"What's a Good Value for R-squared?" What's a Good Value for R-squared? Duke University Accessed August 19, 2015. <http://people.duke.edu/~rnau/rsquared.htm>.

Wood, S.A., Guerry, A.D., Silver, J.M. & Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* 3, 2976; DOI:10.1038/srep02976 (2013).

8 DATA LINKS

NOAA. "2010 C-CAP Land Cover." Digital Coast Data Access Viewer. 2010. Accessed July 24, 2015. <http://coast.noaa.gov/dataviewer/index.html?action=advsearch&qType=in&qFld=id&datareg=1&qVal=1470#>

OpenStreetMap. "Mapzen." Start Where You Are. Accessed August 19, 2015. <https://mapzen.com/data/metro-extracts>

"Parks and Park Features." City of Seattle. Accessed August 19, 2015. <https://data.seattle.gov/Community/Parks-and-Park-Features/rhf2-u5fm>

Seattle Urban Nature Project. "Index of Habitat Data." Index of /ftp/ECScience/Projects/Habitat. September 1, 2002. Accessed July 8, 2015. <http://www.earthcorps.org/ftp/ECScience/Projects/Habitat/>

Appendix A: Python Scripts

A-1: Extracting King County Park Facilities by Type

```
# -*- coding: utf-8 -*-
# -----
# Name: ExtractPointsForNewShapefile_KC.py
# Author: Malena Foster
# Date: 2015-07-24
# Description:
# Extract points by each Location Type and create separate layers for King County
# SeattleParksFeatures_UTM.shp.
#
# -----

# Import arcpy module
import arcpy

# Check out any necessary licenses
arcpy.CheckOutExtension("3D")
arcpy.CheckOutExtension("spatial")

arcpy.env.workspace = "F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\Features\\KingCounty"
in_Layer = "F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\KCParksFacilityPoint_UTM.shp"

typeExtract = ['Access Point (Street)', 'Community Garden',
               'Parking Lot', 'Restroom', 'Swimming Beach', 'Swimming Pool']

outLayers = ['AccessPoint', 'CommGarden', 'Parking', 'Restroom', 'BeachKC', 'PoolKC']

i = 0
for type1 in typeExtract:

    print "layer extract: " + type1
    out_Layer = outLayers[i]
    print "out_Layer: " + out_Layer
    #Extract values for layer
    print "  make feature layer..."
    arcpy.MakeFeatureLayer_management(in_Layer, "temp", "\"F_Type\" = '" + type1 + "'")
    print "  copy feature layer..."
    #arcpy.SelectLayerByAttribute_management("Sitelyr", "NEW_SELECTION", "\"SiteAreaLbl\" = '" + y + "'")
    arcpy.CopyFeatures_management("temp", out_Layer)
    arcpy.Delete_management("temp")

    i = i + 1

print "DONE"
```

A-2: Extracting Seattle Park Facilities by Type

```
# -*- coding: utf-8 -*-
# -----
# Name: ExtractPointsForNewShapefile_Seattle.py
# Author: Malena Foster
# Date: 2015-07-24
# Description:
# Extract points by each Location Type and create separate layers for Seattle
# SeattleParksFeatures_UTM.shp.
#
# -----

# Import arcpy module
import arcpy

# Check out any necessary licenses
arcpy.CheckOutExtension("3D")
arcpy.CheckOutExtension("spatial")

arcpy.env.workspace = "F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\Features"
in_Layer = "F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\SeattleParksFeatures_UTM.shp"

typeExtract = ['Basketball Courts','Beaches',
               'Boat Launches','Ceremonies','Childrens Play Areas','Community Centers',
               'Computer/Media Center',
               'Environmental Learning Centers','Firepits','Fishing','Gardens','Off Leash Areas',
               'Parks','Picnic Sites','Playfields','Pools','Tennis Courts','Viewpoints',
               'Wading Pools','Waterfront']

outLayers = ['BBCourts',
             'Beaches','BoatLaunch','Ceremony','PlayArea','CommCenter','MediaCenter','EnvLearnCntr', 'Firepit', 'Fishing',
             'Gardens', 'OffLeash','Parks','Picnic','Playfields','Pools','Tennis','View','WadingPool','Waterfront']

i = 0
for type1 in typeExtract:

    print "layer extract: " + type1
    out_Layer = outLayers[i]
    print "out_Layer: " +out_Layer
    #Extract values for layer
    print "  make feature layer..."
    arcpy.MakeFeatureLayer_management(in_Layer,"temp","\City_Featu\" = '"+type1+""")
    print "  copy feature layer..."
    #arcpy.SelectLayerByAttribute_management("Sitelyr","NEW_SELECTION","\SiteAreaLbl\" = '" + y + """)
    arcpy.CopyFeatures_management("temp", out_Layer)
```

```

arcpy.Delete_management("temp")

i = i + 1

print "DONE"

```

A-3: Get sum of numbers for each feature and add to parks layer

```

# -*- coding: utf-8 -*-
# -----
# Name: ExtractPointsForAnalysis.py
# Author: Malena Foster
# Date: 2015-08-15
# Description:
# Get summary numbers for each feature in folder and add to CompleteSeattleTUDand PUDs featureclass
#
#
# -----

# Import arcpy module
import arcpy

# Check out any necessary licenses

arcpy.CheckOutExtension("spatial")

outWS ="F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\ParkData\\ParksDataAnalysis.gdb\\"
inputWS = outWS
parksLayer = "F:\\PMP-
GIS\\ArcGIS_Files\\GEOG569\\GIS\\Data\\Seattle_Park_Usage_Capstone\\ParkData\\ParksDataAnalysis.gdb\\Co
mpleteSeattleTUDandPUDs_6"

# OSM only
#fcList = ['BBQ','Field','HardCourt','Parking','SportFac','Toilets']

#list for OSM Data:
#'BBQ','Field','HardCourt',
#'Parking','SportFac','Toilets']
#KC and Seattle
fcList = ['Beaches','BoatLaunch','CommCenter','Fishing','Gardens',
'OffLeash','Picnic','PlayArea','Pools','View','WadingPool']
##

#starting point layer to add the values to
targetLayer = parksLayer

```

```

#i = 0
n=7
for fc in fcList:

    print "Join Features: " + fc

    joinLayer = inputWS+fc
    outLayer=parksLayer+"_"+str(n)

    print "  Get Field Mappings..."

    # Create a new fieldmappings and add the two input feature classes.
    fieldmappings = arcpy.FieldMappings()
    fieldmappings.addTable(targetLayer)
    fieldmappings.addTable(joinLayer)

    #add osm_id field from join featurclass
    #idIndex = fieldmappings.findFieldMapIndex("osm_id")
    # add OBJECTID from join featureclass for Seattle Data
    idIndex = fieldmappings.findFieldMapIndex("InputID")
    fieldmap = fieldmappings.getFieldMap(idIndex)

    # Get the output field's properties as a field object
    field = fieldmap.outputField

    # Rename the field and pass the updated field object back into the field map
    field.name = "cnt_"+fc
    field.aliasName = "cnt_"+fc
    fieldmap.outputField = field

    # Set the merge rule to mean and then replace the old fieldmap in the mappings object
    # with the updated one
    fieldmap.mergeRule = "count"
    fieldmappings.replaceFieldMap(idIndex, fieldmap)

    print "  Perform Spatial Join..."
    dist = 10 #radius search in meters
    arcpy.SpatialJoin_analysis(targetLayer,joinLayer,outLayer,"#","#",fieldmappings,"#",dist)

    targetLayer=outLayer
    #i = i + 1
    n = n + 1

print "DONE"

```

A-4: Clip Tacoma Data Layers

```
# -----  
# Name: automateclip.py  
# Author: Jeremy Forst  
# Date: 2015-08-01  
# Description:  
# Simple clip file script  
#  
# -----  
  
# Import system modules  
import arcpy  
from arcpy import env  
  
# Set workspace  
env.workspace = r"C:\Users\Jeremy\GoogleDriveFW\Capstone_Project-NatCap_Parks\Maps\Map Data"  
  
# Set local variables  
in_features = "Tacoma_KingCo_parks_merged_TUD.shp"  
clip_features = "Tacoma.shp"  
out_feature_class = "Tacoma_TUD.shp"  
xy_tolerance = ""  
  
# Execute Clip  
arcpy.Clip_analysis(in_features, clip_features, out_feature_class, xy_tolerance)
```

Appendix B: EarthCorps Habitat and NOAA C-CAP Aggregations

Table A-1. EarthCorps Habitat Aggregation Groupings

Habitat	Aggregated Group
Landscaped Shrubland	Shrub
Heavy Development	Developed
Shrubland	Shrub
Deciduous Forest	Forest
Conifer Forest	Forest
Conifer Deciduous Mixed Forest	Forest
Grassland	Grass
Landscaped Forest	Forest
Tree Savannah	Forest
Shrub Savannah	Shrub
Conifer Broadleaf Evergreen Mixed Forest	Forest
Riparian Forest	Forest
Palustrine Open Water	Water
Palustrine Scrub-Shrub Wetland	Wetland
Riverine Unconsolidated Substrate	Water
Landscaped Grassland	Grass
Palustrine Forested Wetland	Wetland
Medium Development	Developed
Landscaped Tree Savannah	Forest
Riverine Consolidated Substrate	Water
Palustrine Aquatic Bed	Water
Light Development	Developed
Herbaceous Row Crop	Crop
Riverine Tidal	Water
Cliff	Rock
Palustrine Emergent Wetland	Wetland
Sparsely Vegetated Habitat	Grass
Deciduous Broadleaf Evergreen Mixed Forest	Forest
Beach and Dune	Beach
Orchard or Vineyard	Crop
Broadleaf Evergreen Forest	Forest
Rock and Talus	Rock
Lacustrine	Water

Table A-2. NOAA C-CAP Land Cover Aggregation

NOAA Class Name	Aggregated Group
Bare Land	Bare Land
Cultivated	Open Space
Deciduous Forest	Forest
Developed Open Space	Open Space
Estuarine Aquatic Bed	Water
Estuarine Emergent Wetland	Wetland
Estuarine Scrub/Shrub Wetland	Wetland
Evergreen Forest	Forest
Grassland	Open Space
High Intensity Developed	Developed
Low Intensity Developed	Developed
Medium Intensity Developed	Developed
Mixed Forest	Forest
Palustrine Aquatic Bed	Water
Palustrine Emergent Wetland	Wetland
Palustrine Forested Wetland	Wetland
Palustrine Scrub/Shrub Wetland	Wetland
Pasture/Hay	Pasture
Scrub/Shrub	Scrub/Shrub
Unconsolidated Shore	Beach/Shore
Water	Water