Urban Areas and Avian Diversity: Using Citizen Collected Data to Explore Green Spaces

by

Ephriam Joseph Daniels

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

August 2019

This thesis is dedicated to Yi-Hsuan Chu who encouraged me to pursue my dreams and to my family for their loving support

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

I am grateful to USC faculty for the direction I needed and my other faculty who assisted me when I needed it. This is especially true with my thesis committee members; Dr. Bernstein, for insight in developing my thesis, Dr. Wu in organizing and manipulation my database and Dr. Marx for biological insight when working with GIS. I am also grateful for Dr. Quang's lab at Academia Sinica for teaching me R-studio and statistical applications.

List of Abbreviations

BLI	Birdlife International
GIS	Geographic Information System
JSI	Jaccard Similarity Index
MIT	Ministry of Interior of Taiwan
NSLC	National Survey and Land Center
NDVI	Normalized Difference Vegetation Index
ODBI	Open Database Connectivity
SSI	Spatial Sciences Institute
USC	University of Southern California
UHI	Urban Heat Island
VGI	Volunteer Generated Information

Abstract

Urban development is expanding in today's world, and the impacts on humans and the environment are strained within these modern cityscapes. The Urban Heat Island phenomena and habitat loss have raised concerns about the future of many the ecological health of many metropolitan areas. Due to these concerns, cities have taken steps to reduce the negative impacts on the urban environment with the use of green spaces. On the Island of Taiwan, the municipality of Taipei is one metropolitan area that has experienced dramatic urban growth. While multiple studies have investigated the avian diversity of Taipei's green spaces, most studies have used traditional data collecting methods. These surveys are financially taxing and time-consuming, which can limit the volume of recorded events. In this study, Volunteer Generated Information (VGI) and Geospatial Information Systems (GIS) was used to determine the biodiversity, richness, and species composition of 25 green spaces selected by data-driven selection process within Taipei from 2016 to 2018. The eBird dataset and multiple indexes served as indicators of ecological health allowed for monitoring of green spaces. This study determined that there are relationships between biodiversity, richness and species composition of green spaces within Taipei. However, specific site’s species compositions in VGI showed weak links between the richness of the green space. VGI datasets and GIS could enable a cost-effective way to monitor a city's green spaces effectively in the future.

Chapter 1 Introduction

Human migration from rural regions into urban areas has produced multiple negative implications to both humans and the animals that inhabit these urban areas. Adverse outcomes are caused by two large factors; urban shift and the Urban Heat Island (UHI) phenomena. To combat these factors, many cities have created or improved designated green spaces. Studies have investigated specific green spaces in the city of Taipei Taiwan. However, there were large constraints on the data that was collected (e.g., time, locations, etc.). Citizen-collected data can be used to explore multiple areas that are less limited by these constraints while still having higher avian counts than previous studies. Therefore, using these types of datasets can increase the number of selected sites based on the dataset and not the pre-selected sites.

1.1. Urbanization and Green Spaces

The pressures of urbanization will increase with time and will have significant impacts on urban environments. Currently, 54% of the human population lives in urban areas; this number is predicted to grow to 60% in the next 30 years (The World Bank, 2018; United Nations, 2011). This urban migration will need to be managed by city planners to promote a healthy city. These city planners must focus not only on social aspects when designing and improving urban areas, but also environmental ones. Urban migrations impact local ecosystems and displace several native species (Barbosa et al. 2007). That said, the urban shift is not the only factor that causes environmental displacement.

As urbanization increases globally, concerns arise due to the shift from rural regions to more densely populated human-made environments. One concern with this urban shift is the UHI phenomenon. One of the phenomenon's traits is an increase in temperature within urban areas, which can have adverse effects on both humans and the environment. The UHI phenomenon increases health risks and wildlife displacement due to the increased temperature (Akbari & Kolokotsa, 2016). Awareness and concerns about this phenomenon have led to the improved urban design and designation of green spaces.

Green spaces and designated vegetated areas have been used by city planners to address the effects of urban shift and the UHI phenomenon. Green spaces and the diverse vegetation provide benefits both to humans and the environment (Wolch, Byrne, & Newell, 2014). But not all green spaces are created equal. Many studies have found that the spatial distribution of green spaces (islands vs. networks) and the overall site quality, can impact positively the biodiversity of the green space (Aida et al., 2016; Chamberlain et al., 2007; Imai & Nakashizuka, 2010). Because of this effect of green spaces, many cities have started implantation more green spaces or improving existing ones within their administrative boundaries.

1.2. Avian Species in the Green Spaces

Many studies have used avian species as indicators of an ecosystem's health (Melles, Glenn, & Martin, 2003). One of the advantages of using avian species versus other organisms is the ability to differentiate species by phenotypical traits and calls easily. Understanding species richness and diversity within a specific region allows researchers and city planners to interpret the level biological of stability and success of the green space post-implementation. Multiple studies have linked the ecological health of green spaces to the richness and biodiversity of avian species (Shih, 2010; Tajima, 2003). They have also noted that the area of green spaces may play a smaller role compared to the biodiversity that is within its boundaries.

In recent years, research has been conducted on avian species and their distribution in urban areas. Findings concluded that diversity in green spaces within urban areas dramatically increases the biodiversity of the municipality and its surroundings (Chamberlain et al., 2007;

Shih, 2018; Strohbach, Lerman, & Warren, 2013). This increase of avian biodiversity in green spaces has been studied in the city of Taipei by Dr. Shih (Shih, 2018). However, the Shih study had low recorded sightings and a limitation of the study period. Because of these limitations, a misrepresentation of the green spaces could be observed. By using geospatial information systems (GIS) selection tools and citizen collection datasets, this study explored the relationship between biodiversity and green spaces of Taipei. Because of the dataset used and the method for site selection allowed for a more robust observational dataset for a more extended period.

1.3. GIS and VGI Data within Taipei

With the increased adoption of geospatial technologies, researchers can more easily identify and explore green spaces. The ability allowed by GIS applications allows for more reliable and quicker collected datasets than traditional monitoring methods (Silvertown 2009). This improvement in methods is partially due to the increase in handheld devices and intuitive collection applications. When compared with previous studies, the use of volunteer-generated information (VGI) data has proved useful when working with conservation and/or monitoring efforts (Chandler et al., 2017). These tools can be used in cities around the world.

The municipality of Taipei is located within the Taipei Basin which is surrounded by mountains in the Island of Taiwan in East Asia (Figure 1). Most populated parts of the municipality are alongside the Tamsui River. While the center of the city is at low elevation, much of the municipality is within a steep mountain range. The municipality of Taipei is in the subtropical climate. The temperature in Taipei is warm, with precipitation moderate to high all year round (Chang, Li, and Chang 2007). From June to October Taiwan experience typhoons, tropical cyclones that occur in the Pacific Ocean. Because of the high levels of precipitation and being in a subtropical region, Taipei is conducive to having a high amount of vegetation. Even with high amounts of vegetation, the municipality has changed over the years.

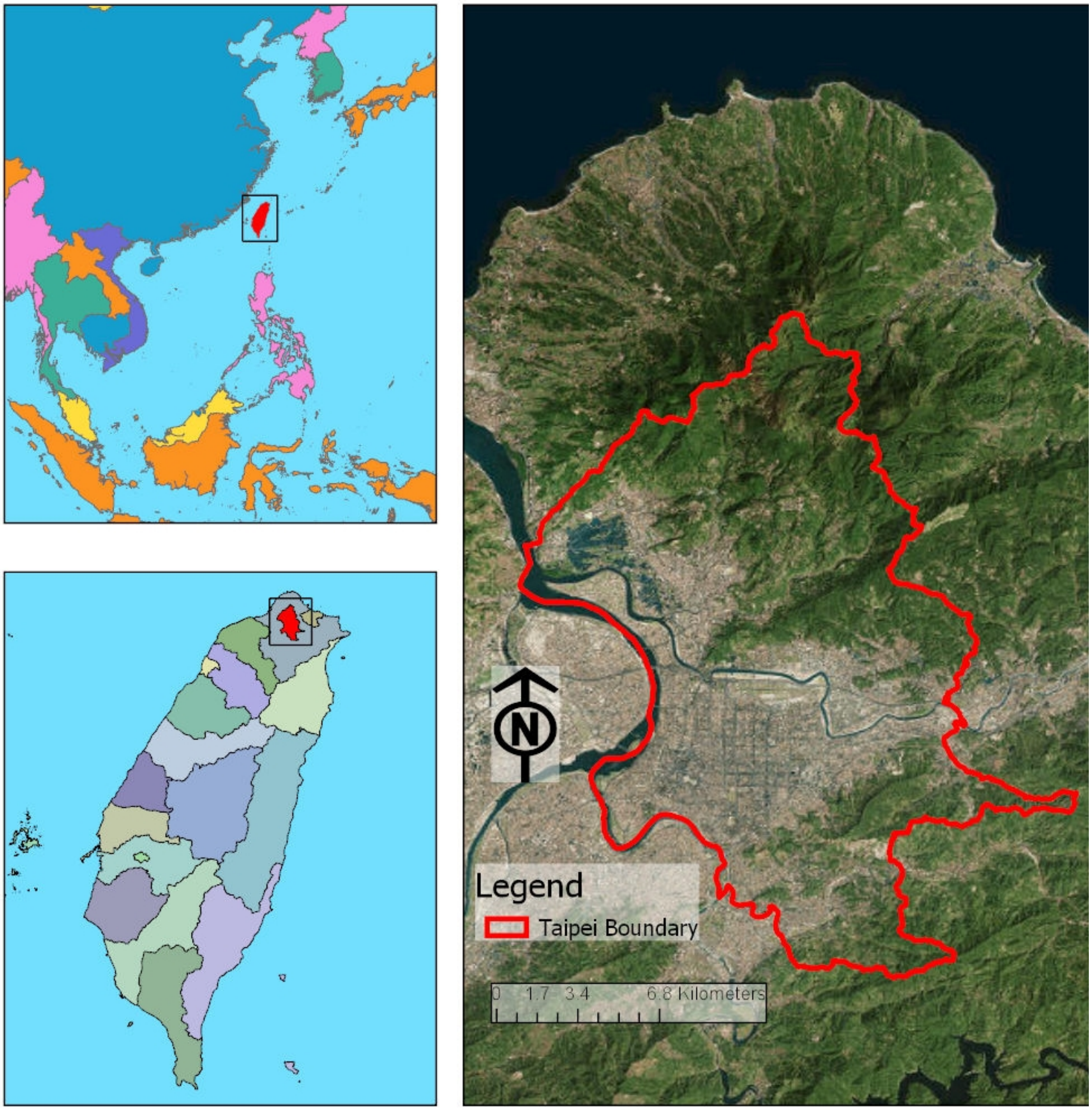


Figure 1: Administrative boundary of the municipality of Taipei.

Dating back to the 18th century, Taipei has been the capital of Taiwan and is currently home to over 2.7 million people with over 5 million people that commute between its administrative borders (Shih 2010). The dense population has resulted in multiple governmental concerns. One of these concerns is the effects population density on the ecological health. The loss of native ecological habitats has led to limited resources for animals within cities (Goddard, Dougill, and Benton 2010). This limitation is compounded with the increase of multi-story buildings which have been linked to increasing the average city temperature which leads to the displacement of multiple native species (Tratalos et al. 2007). Because of these concerns, the municipality of Taipei has taken steps to improve the health and wellbeing of its ecosystems with the use of green spaces. Green spaces have been used to promote biodiversity in Taipei, yet few studies focus on showing their effectiveness.

There have been significant changes in improved urban planning policies over the years in the municipality of Taipei that have been beneficial for native species. These policies have been used to increase the construction of green spaces as well as increase the diversification of vegetation of existing green space (Weng, Lu, and Schubring 2004). This increase in green spaces can improve multiple native faunae. Taiwan is home to a diverse number of avian species. According to Birdlife International (BLI), Taipei is listed as having 342 species within its municipality (BirdLife 2018). While there are a large number of species that can be found within Taipei, the biodiversity of any single location can be high depending on the quality of the site (Goddard, Dougill, and Benton 2010). This high species count in part is due to Taipei location which is found at a low altitude with mountainous and aquatic features (Huang and Chan 2014).

1.4. Objectives

The objective of this study was to use geographic information science and volunteer-generated information datasets to monitor biodiversity and ecological richness in green spaces in the city of Taipei, Taiwan. While past studies about the relationships between green spaces and avian distribution have been conducted in Taipei, none have used VGI data. The VGI dataset consists of the higher volume of avian sightings and allows for a more supported conclusion of the ecological health of Taipei's green spaces. This high volume of records would support the study to understand the ecological health of the data selected green spaces within the municipality of Taipei through multiple years.

There are two specific research objectives in this study. The first objective is to examine how biodiversity and richness varies across green spaces in Taipei. Biodiversity is defined here as the abundance of species in selected areas and not by the dominant species. For this both the Shannon-Weiner Index and rarefication analysis on each of the selected sites. The study compared the levels of biodiversity and richness between different sites. The study also explored the relationship the green spaces have over multiple years. From January 2016 to December 2018 each green space's average biodiversity was determined. The rarefication analysis will further determine the most optimal sample size needed to be collected for green space. These measurements will allow an understanding of the ecological health of each location and optimal sample size for VGI datasets.

The second objective of this study is to determine if there are links between selected green space's biodiversity, richness, and species sites composition. This was further differentiated between resident vs. migratory species, as many sites may have a stronger abundancy of species depending on the species residency status. The understanding of the species residency statuses and how they are distributed between the green spaces better defines a link between specific green spaces to species composition. To further support this link, species site composition of all selected green spaces was analyzed using the Jaccard Similarity Index

(JSI). By exploring these links, this study will provide an understanding and further support the final objective.

The results of the study can be used to not only identify the quality of differing green spaces based on their avian biodiversity but will also allow a deeper understanding of which types of avian species compositions are affecting specific green spaces locations inside Taipei. Both sub-questions will support the finding in the primary objective of the study.

Chapter 2 Background

This chapter describes the background information related to this study and focuses on four subjects: Urban shift and the UHI phenomenon, green space, and urban vegetation, Taipei green space and their vegetation endeavors and finally citizen-collected data and GIS monitoring.

2.1. Urban Shift and the UHI Phenomenon

Much of the global population is found in urban centers. The U.N. predicts that over 60% of the earth's population will live in cities by 2050 (United Nations 2011). This endeavors to create urban center needs land and large amounts of resources in concentrated areas. Urban development influences human and non-human life within the population center.

The human population is migrating from rural areas to highly populated urban areas. This change in population placement has had a pervasive ecological cost. While most rural housing is spread out over a broader region, cities condense their housing area. This urban shift effects various species as it narrows the ecosystems they can have and limits the available natural resources while displaces multiple species that were originally native to a formerly rural area. This phenomenon will degrade the quality and function of an ecosystem. This degradation is only compounded by other effects caused by the UHI phenomenon (Shiflett et al. 2017).

The construction in urban development due to population migration significantly increases the amount of human-made material in a concentrated area, and many of these materials (asphalt, cement, metals) are quick to collect solar radiation (Chen et al. 2006). Because of the low specific heat capacity that is found in human-made materials, many urban constructions can collect so much solar radiation they increase the surrounding temperature of the area. The UHI has led to multiple issues including a decrease in biodiversity and multiple health concerns within larger cities (Guo et al. 2015; Tan et al. 2010; Akbari and Kolokotsa 2016). Because urban areas are prone to multiple factors related to UHI and urban development, city planners have taken steps to reduce the UHI effects.

2.2. Green Spaces and Urban Vegetation

The United States Environmental Protection Agency (US EPA) (2018) defines open spaces as a piece of land that is undeveloped and accessible to the public. Of all the categories open spaces, green spaces have been shown to have a cooling effect on both local and surrounding regions which has been documented to reducing the effect of UHI (Estoque, Murayama, and Myint 2017). Greens spaces defined in this study are open spaces designed to be vegetated for human, and non-humans use. Much research has been done on the health of humans and the environment with the creations of these spaces (Barbosa et al. 2007; Fuller et al. 2007; Wolch, Byrne, and Newell 2014). One of the findings is that green spaces because of their vegetation produce shade, which reduces stored thermal energy in the surrounding areas (Li et al. 2011). Because of this finding, remote sensing has been used to locate areas of interest. Datasets that use thermal sensors have been linked to areas of vegetation (Tan et al. 2010). The Normalized Difference Vegetation Index (NDVI) has been used in multiple studies to locate areas suspected of reducing temperatures because of thermal storage (Li et al. 2011). This cooling effect has been seen to reduce the power needed to cool electronics and reduce the amount to ozone, which impart reduces cardiovascular disease (Akbari and Kolokotsa 2016). While there have been direct links between human health and financial benefits of green spaces, there are also links connecting green spaces to ecological health.

Richness and biodiversity are two defined terms that have been used to determine a location's ecological health. Richness is defined as the number of individual species in a specific ecological location (James and Rathbun 1981). This term is calculated by the sum of all species in a specific location. While there is a single method for determining the richness, there are multiple indexes used for biodiversity. Biodiversity is defined as the number of individuals and a variety of species in an ecological community (Chao et al. 2006). While there is a single definition, the term is often broken into two different methods depending on a study's focus: Simpson Index and Shannon-Wiener Index. The Simpson Index is weighted by dominant species, which is used to measure the degree of concentration when individuals are selected for specific groups (Oksanen 2016). The Shannon-Wiener Index focuses on the abundance which quantifies the uncertainty of the predicted species in an area (Tramer 1969). This index was used in this study as it works on the abundance of a species in a specific site. Understanding both the richness and biodiversity of the term allows for insight into green spaces.

Green space has been shown to improve their effectiveness to an environment's biodiversity, as they enable multiple biological niches that were not previously present (Goddard, Dougill, and Benton 2010). Greens paces have been linked to improved health in both people and the biophysical environment (Tan et al. 2010; Wolch, Byrne, and Newell 2014). Green spaces with high levels of biodiversity have affected multiple species as these organisms fill new ecological niches (Matsuba, Nishijima, and Katoh 2016). While green spaces to increase the amount of biodiversity in an area, there is a limit to that diversity. Researchers determined that with the rarefaction test, different sites have a maximum richness threshold in avian diversity depending on the variables of the site (James and Rathbun 1981). This finding gives support to the idea that each site's features determine different environmental health as levels of both richness and biodiversity are not shown to be equal.

One standard indicator used for environmental health studies is the diversity and abundance of avian species. Studies have used avian species to monitor the health of specific environments (Blair 1999). Avian species are apex organisms that are relatively easily to identify compared to other organisms and have been linked to the biodiversity of the surrounding areas (Kong et al. 2010). This link was further studied when examining green space relations to the residency of differing avian species (Matsuba, Nishijima, Katoh 2016). As different species need different requirements in their habitat, trends start to emerge depending on the species residency. While there are overlapping areas, migratory and resident avian species have been more conducive to specific areas (Fontana et al. 2011). This spatial pattern is further limited in a cityscape's green spaces where natural vegetation confined and vary significantly (Shih 2010). Because of a green space's limitation, multiple species compositions have been shown to highlight areas that are conducive for higher levels of biodiversity. The JSI has been used to not only to determine these sites but show sites relation to varying levels of richness (James and Rathbun 1981). Because of avian and their biological status are classified as higher-level organism, it is easy to discern and link them to green spaces.

2.3. Taipei Green Space and Urban Vegetation Endeavors

Taipei is in a sub-tropical climate that is conducive to yearlong vegetation growth. With most of the city inside the Taipei basin, the location has allowed the city to have a relatively large amount of vegetation and a diversity of animal species. When Taiwan started to industrialize in the mid-1970s, the number of green spaces was reduced (Taipei DUD, 2009). In recent years, Taipei has seen a rise in its urban population and has become one of the most densely populated cities in the world (Shih 2018). Started in 1992 with the Taipei

Comprehensive Urban Plan, Taipei city planners first focused on urban development with little emphasis on the biophysical environment (Shih 2010).

With concerns about UHI phenomena and intent on improving the city's image, Taipei began using planning infrastructure changes to improve urban ecosystems in the early 1990s (Shih 2010). It was in 1999 that development policies included the increase in green spaces. The emphasis in urban vegetation was further expanded with the creation of the Green Master Plan for Taipei (Huang and Chan 2014). After the establishment of the Green Master Plan, Taipei implemented the Beautiful Taipei Policies to set specifics policies for land improvement that included increasing the number of green spaces and urban vegetation locations in 2010 (EPA, ROC 2018). Because of these policies, the city increased and improved vegetative diversity many green spaces within its administrative borders. Currently, the city of Taipei's Department of Urban Development contains 281 units of designated green spaces and over 80,000 trees planted on roadsides (Taipei DUD 2009). This number does not account for how the city, as it breaks larger units into multiple parcels of green spaces. While studies have been used to predict the best placements for green spaces, there is little work on analyzing their efficacy with respect to biodiversity after they have been created. This exploration can be done with the use of avian species and their compositions as seen in earlier studies (Shih 2010). With this information, one can investigate how effective green spaces are to the ecological health to where they have been implemented in Taipei.

2.4. Citizen Collected Data and GIS Monitoring

Citizen-collected data has been gathered and used for hundreds of years. Charles Darwin was an unpaid companion to Captain Robert Fitzroy when he traveled on the Beagle collecting data on wildlife (Silvertown 2009). With the emergence of new technologies, data collection has become much easier and more accessible. As a result, the volume of citizen-collected data has grown immensely (Silvertown 2009). Many organizations have committed to storing and maintaining large datasets including citizen-collected data and making them accessible to the general public.

New technologies allow users to achieve a clearer understanding of ecological information. GIS has been used in avian studies to find patterns of migration and habitat ranges (Bouten et al. 2013; Fink et al. 2011). Researchers have also been able to use geospatial analysis to understand urban green spaces and the correlations it has with species biodiversity (Matsuba, Nishijima, and Katoh 2016; Wiens et al. 2009). In recent years, there has been an increasing amount of research using GIS with open-source databases. The availability of datasets has only increased with the inclusion of VGI.

There has been an increased use of VGI databases. The accessibility and increased usage of VGI datasets have been shown improved their reliability (Silvertown 2009; Conrad and Hilchey 2011; McKinley et al. 2017). This reliability is primarily supported by the number of recorded events that VGI databases can produce. VGI database allows for a larger number of moderators, which in turn gives more opportunities to verify large dataset quickly. These datasets give public access to record and increase their ability to analyze geospatial data. They also encourage users to add more data to datasets as many VGI databases are simplified for quick data collection. The dataset grows and improves with every additional record. This increase in shared data also improves the accuracy of a dataset. These datasets increase on both the temporal and spatial scale of a study which could not be conducted by traditional means, which requires a team of researchers that could only collect data for a limited amount of time (McKinley et al. 2017).

Chapter 3 Methods

The study was conducted within the urban areas of the city of Taipei and focused on the distributions of avian diversity and richness within green spaces. The study was conducted in three analytical phases; (1) primary data collection/exploration, (2) data preparation, and (3) analysis (Figure 2). Each phase was vital to the study as each was dependent on the prior stage.

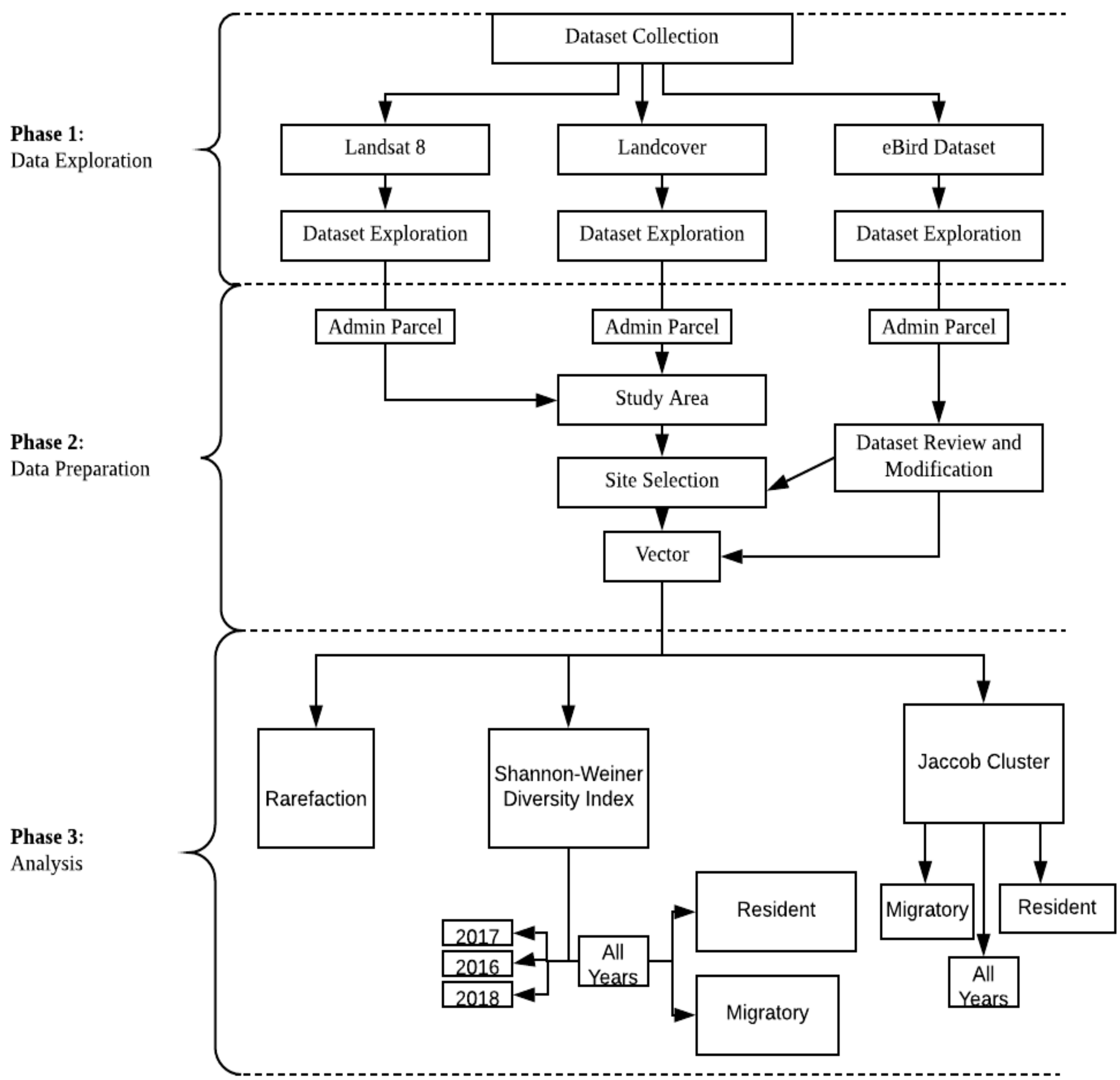


Figure 2: The workflow diagram for the methods used in this study.

3.1. Data Exploration and Management

In the first phase, datasets were collected from various sources (Table 1). Both QuantumGIS (QGIS) and PostgreSQL applications were used to explore and store different datasets. Because of availability and the user-friendly design, ArcGIS Pro was used for exporting final map/figures. Five datasets were collected for this study; eBird, BirdLife International, Taipei administrative parcel, Taipei landcover, and Landsat 8 imagery.

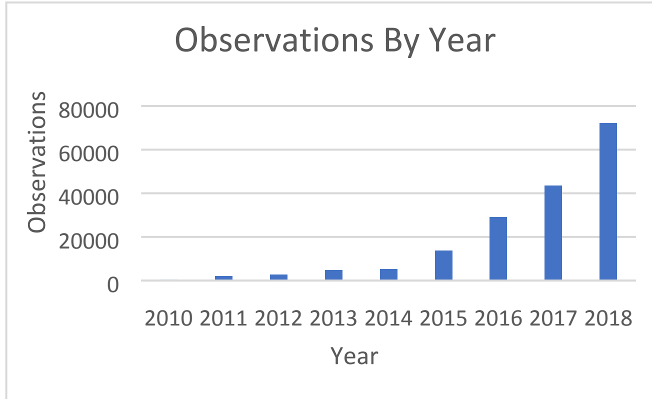
Dataset	Type	Description
EBird Dataset	Vector – Point Original Txt.	EBird dataset collected from March 2010 to 2018 https://ebird.org/
Birdlife International	Table	Information on species names in municipality of Taipei Updated 2018 http://datazone.birdlife.org/home
Taipei Administrative Parcel	Vector – Polygon Original Shp.	Administrative borders of Taipei region. Updated 2018 https://data.gov.tw/
Taipei Landcover	Vector - Polygon	Landcover of Taipei City Updated 2018 http://www.nlsc.gov.tw
Landsat 8 image	GeoTIFF	Landsat NDVI 30-meter cells of 2018 from Climate Engine https://app.climateengine.org/

Table 1: Datasets used in study

This study utilized the eBird datasets, which is a collection of citizen-collected information on avian species around the world. This original dataset utilized eight years data which was recorded from January 2010 to December 2018. The dataset was selected for the area within the municipality of Taipei using the eBird portal. The original dataset was a text file and was first added into QGIS to make the dataset spatial by converting each record into points using the XY coordinates. The data was then converted from a geographic coordinate system to a projected coordinate system in Taiwan Datum 1997 (TWD97). The projection used throughout the study was TWD97. The dataset was transferred into PostgreSQL for pre-analysis.

EBird datasets have one of the most extensive collections of VGI on avian species worldwide. While the dataset is large, there are still variables that need to be understood to be used accurately for this study. One of the concerns with the citizen-collected dataset was the quality control that may vary between users. The dataset allows users to investigate the collector, collection method, and if the specific recorded had been peer reviewed. This process to investigate the eBird dataset was done through PostgreSQL, as the application was able to handle the entire dataset efficiently. SQL script was made to find the sum of yearly observations, the number of protocols used throughout the datasets (Appendix A, B). The reduction of anomalies of the eBird dataset and yearly diversity were counted and refined using PostgreSQL (Table 2). The data used in the study was from 2016 to 2018, as earlier years did not have sufficient recorded events. The final data selection of the study was further supported, as the richness of avian inside Taipei were starting to level from 2016. While the eBird dataset was suitable for this study, there are limiting factors.

eBird Dataset		
Total Counts	Richness	Year
462	103	2010
1915	143	2011
2743	164	2012
4691	182	2013
5100	206	2014
13721	253	2015
29129	294	2016
43383	324	2017
72228	311	2018



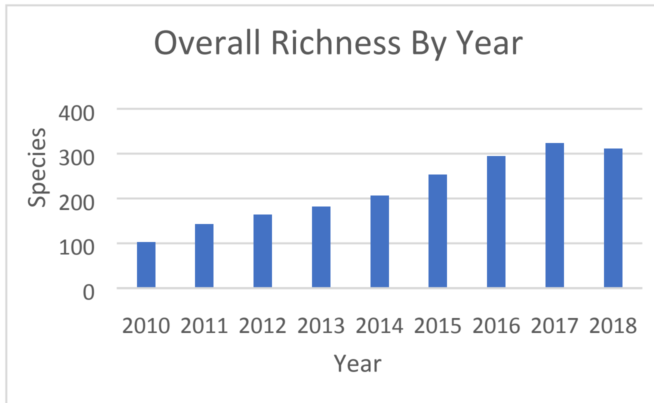


Figure 3: The explorative analysis of the eBird dataset. (Left) The recorded diversity of avian species in Taipei from 2010-2018 (Top Right) Avian species recorded of Taipei from eBird dataset (Bottom Right) Richness of avian species found in eBird dataset

While the eBird dataset has a high number of recorded events, it has some drawbacks. There was potential bias in how humans recorded specific areas unrelated to avian distribution. Multiple factors can cause these biases. The first factor is that the point of the collection is not always the location of the bird, but rather the position of the collector. With this designation of location, the distance between the collector and the bird were typically ranged between 0 to 30 meters. This protocol is marked in eBird data as P25. The second factor is with high counts but with low time expanded for observations. Some of these cases were a result of different methods of collection and not collector's error. Because of this multimethod dataset, records that were significantly different for the rest of the dataset were omitted unless the appropriate time for the collection was given. The final factor that can affect distribution is errors in collectors recording species names. EBird dataset allows users to add genius names without recording the species, because of this richness and biodiversity may be artificially heightened because of the lack of user precision. This correction of nomenclature was corrected using the Birdlife International (BLI) dataset as it would have an up-to-date and standard species checklist.

The Birdlife International dataset verifies of scientific names, residency, species status. BLI is an organization that categorized avian distribution throughout the world. The dataset also contained threatened status which was used to explore if there are any anomalies with the eBird dataset. The BLI dataset was selected to represented avian species presence in the municipality of Taipei. Because of this BLI dataset, only 117 resident and 225 migratory species were selected for this study.

To define a study area, an administrative parcel dataset was used. This dataset was collected from the Ministry of Interior of Taiwan (MIT) online portal. This dataset gave the administrative boundaries of all the counties and municipalities within the Island of Taiwan. This dataset was suitable when defining the study area as it contains the municipality of Taipei's boundary. A text file contained parcel's information and was obtained through the MIT web portal. A join between the parcel shapefile and the text file was be done in QGIS. The dataset was then be added to the geodatabase in PostgreSQL for future processing.

To define the study sites, an administrative landcover dataset was used. This dataset was collected from the Taipei's National Land and Survey Center (Figure 4). This landcover dataset used was a collection of 38 individual tiles of Taipei and the surrounding areas taken from 2018 with the resolution of 5:000. Because of inconsistencies in dataset polygons, only 37 tiles were selected for the study.

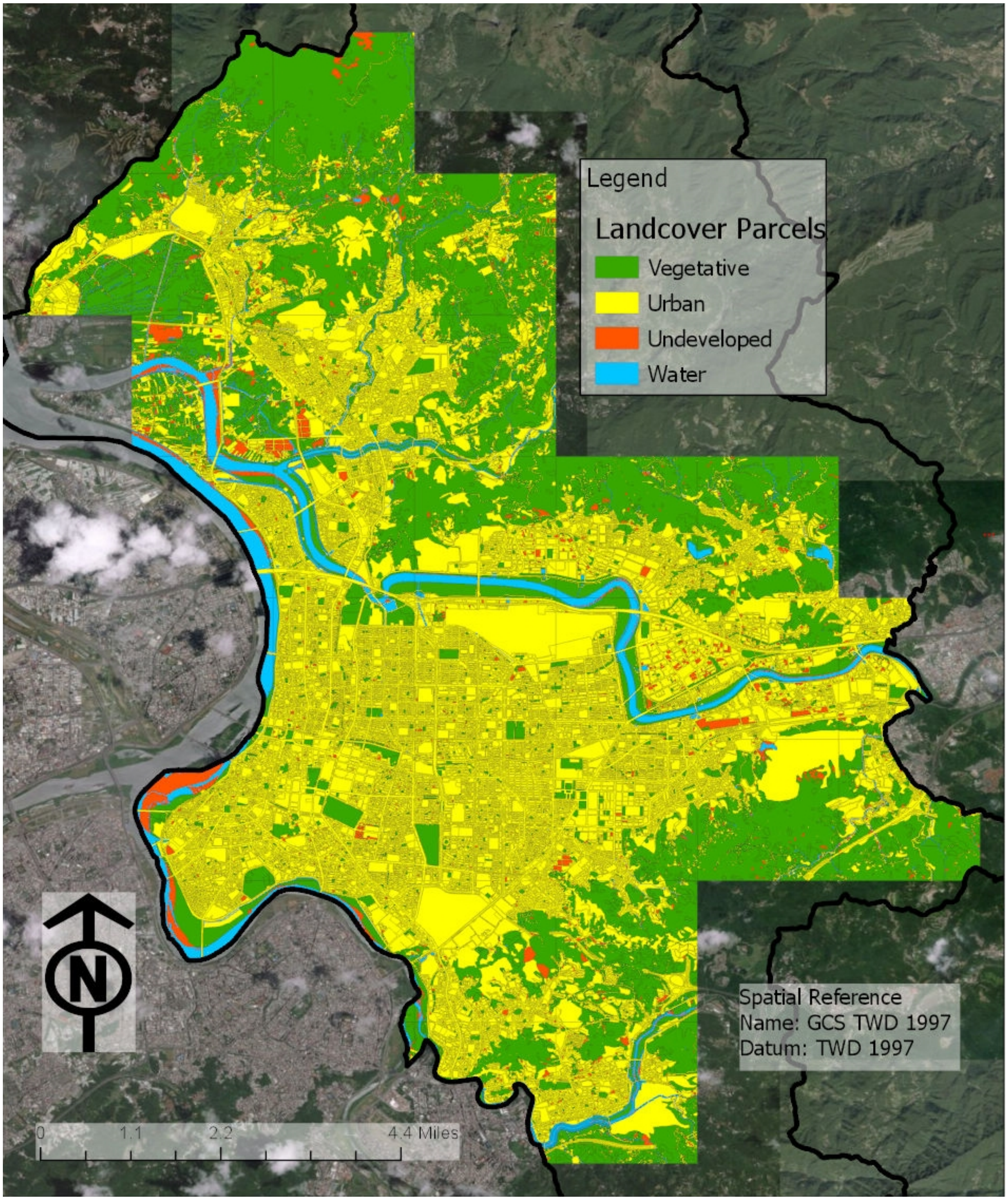


Figure 4: NLSC datasets representing Landcover of Taipei of 2018

Landsat 8 Level 1 data was used in the study to determine the extent of urban areas. The dataset was be collected from Climate Engine. Climate Engine is a web-based application that allows for the procurement of opensource high-resolution global remote sensing datasets. The dataset has a 30-meter spatial resolution NDVI. As the parcel dataset was last updated in 2018, the Landsat image was used a mosaic average from January 2018 to December 2018. The dates of capture ensure quality control for defining the study area as it ensured the latest images of urban expansion.

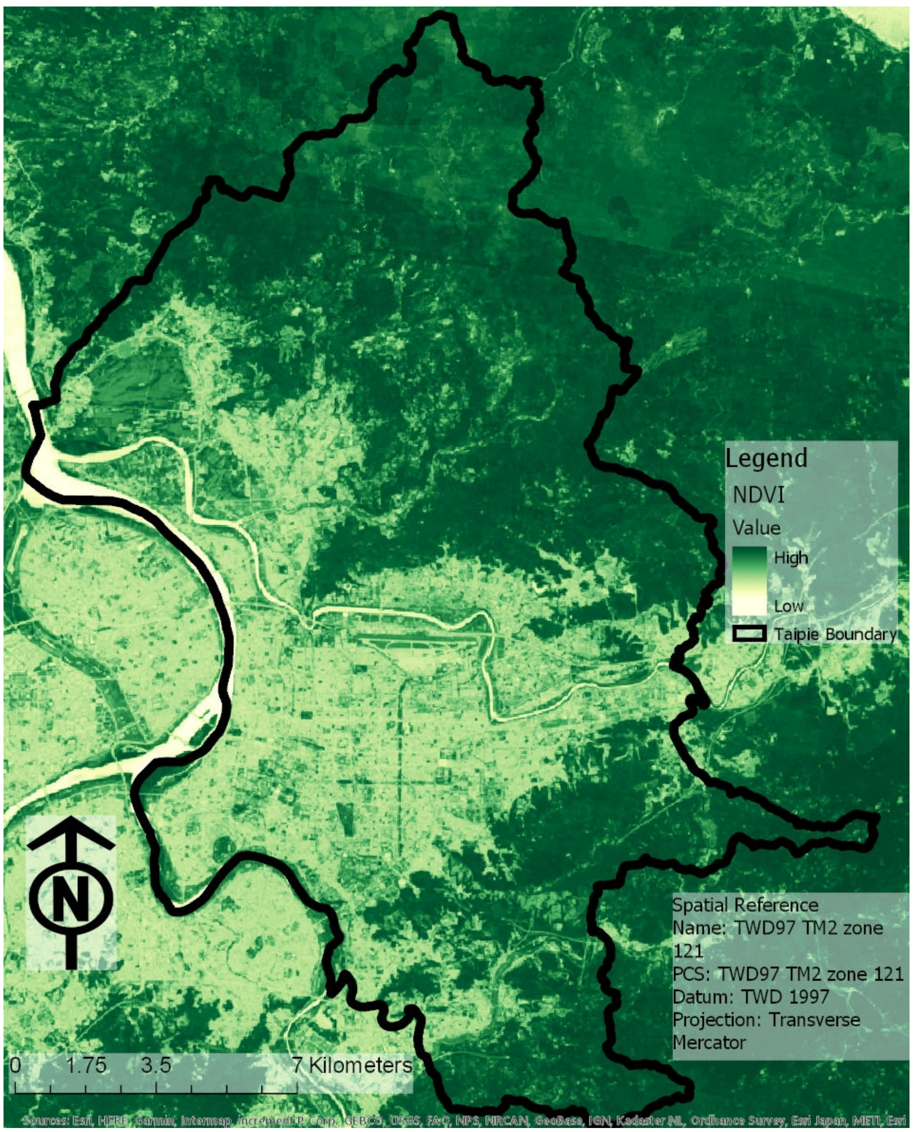


Figure 5: Landsat 8 NDVI of Taipei Area

3.2. Data Preparation

The second phase is the data preparation where datasets were created and manipulated. In this phase, there was an emphasis on organizing the newly created datasets (e.g., sites table, urban areas) in a geodatabase (Figure 6). A geodatabase was created, connected and updated between QGIS, PostgreSQL, and R-Studio for the ease. EBird and site selected tables will be created so there is no many to many relationships within the database. Landsat, Taiwan administration, selected sites, and eBird datasets will be connected with their geometry under the TWD97 projection. Much of the data processing was completed in both Terrset and PostgreSQL. QGIS was used as an intermediate software. Final selected sites created in QGIS was compiled into the geodatabase which was stored in PostgreSQL for ease of data manipulation and organization.

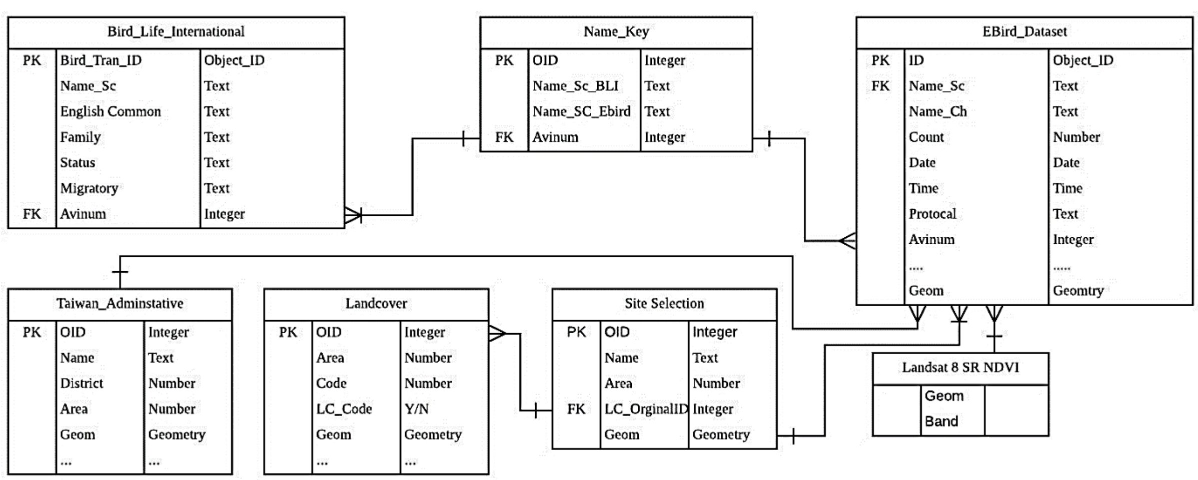


Figure 6: The Avian Database

The process of defining a study area was conducted with the Taipei administrative parcel dataset. Selecting only the Taipei municipality out of the entire country of Taiwan allowed for a clear definition of the municipality's borders. The urban area in which the selected sites will be selected will be extracted using Terrset and QGIS.

3.2.1. Determining All Green Spaces

The selected sites were inclusive to only the urban areas of the city of Taipei to represent the populated areas of Taipei. Urban areas were defined using Terrset and QGIS. This process was conducted using a Landsat 8 Level I NDVI and the Taipei parcel dataset. The Landsat 8 Level I's NDVI was used to find the N-value that were between .082959 and .31047. This value was determined with the use of the histogram tool within Terrset. The N-value peaked at three separate occasions: vegetated areas, urban areas, and water (Appendix C). Using the Reclass tool in Terrset a new image was created, where the values -1 to .082959 were classified as water, value .082959 to .31047 were urban areas, and values greater the .31047 were vegetative areas. Before exiting Terrset, the new image was converted into a vector file. This new vector dataset was imported into QGIS.

In QGIS the newly created urban area dataset was used to determine all the green spaces within the municipality. The urban dataset had a 50-meter buffer added to incorporating any green spaces that could have been on a border area (Figure 7). This buffer was selected as some larger parks were not originally selected even if they were found in the center of the city. This new dataset was used to initially define the urban areas by creating a new border shapefile within QGIS. The parcels that are within the new border shapefile will be qualified for the new study area. The study area covers a total of 271 square kilometers and has 3,100 parcels of landcover were designated green spaces that were applicable for the study area (Figure 7). The undeveloped land was not designated as possible green space as their attributes varied significantly.

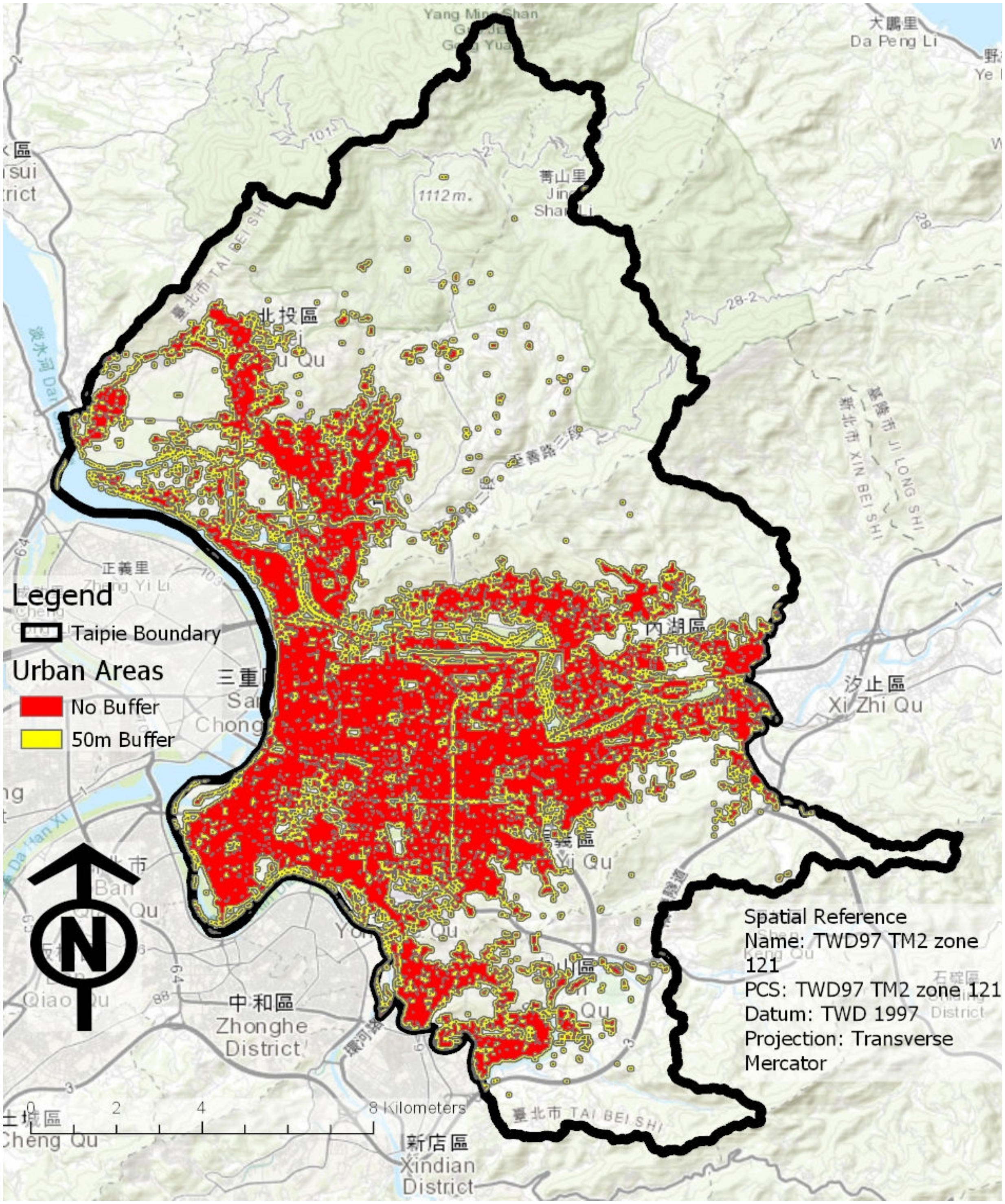


Figure 7: Urban areas within Taipei with 50m buffer

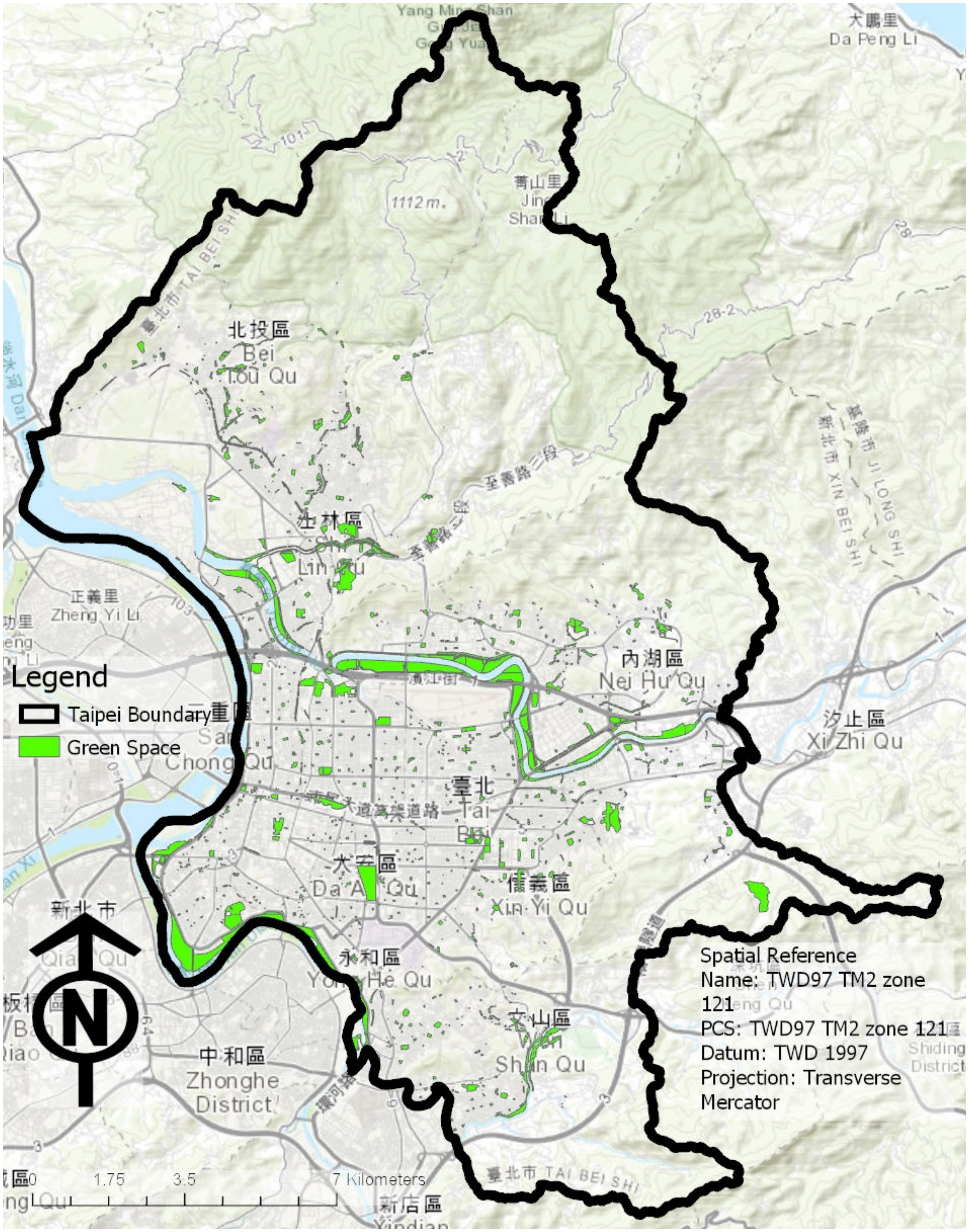


Figure 8: All possible green space found within Taipei from landcover dataset

Within the municipality of Taipei, there was a total of 3,100 designated green spaces identified from Section 3.2.2. Using PostgreSQL, 25 specific sites within the identified green spaces were identified for further analysis. This site selection procedure was completed in multiple steps using a QGIS's database manger and QGIS editing tools. However, all SQL script was tested in PostgreSQL before use in QGIS (Appendix B) .

The first step was to identify the identified green parcels that are also classified as green spaces within the Taipei landcover dataset. The attributes of the landcover datasets, values with an ID value between 700 series were classified as green spaces by the NLSC. The use of SQL statements enabled the selection of the attribute series which corresponds with green areas in the dataset. As avian species can have a broader range than the exact location in which the bird was identified, a 300-meter buffer was created around selected sites for the avian count (Shih 2018). This script was combined with an intersect SQL statement to determine the number of recorded avian counts inside the newly acquired parcel zones. Only green spaces that had equal to or greater than 100 avian observations inside their boundaries were selected. The was exported as a vector dataset in QGIS through the database management tool in QGIS.

Initially, the top 300 were selected using the SQL script. From those 300 parcels, the merge commands were conducted in QGIS to connect the parcels that belong to a single site. The Open Street Map background was used to determine its borders. As the landcover dataset separates green spaces by NLSC's designated parcel, it can mispresent a larger site. This misrepresentation is because some of the sites are a combination of multiple parcels. This process changed 300 parcels into 257 potential sites. This new dataset was uploaded into PostgreSQL under the table labeled sites. Using the SQL script to find the 300 parcels was modified, selecting from the new sites table with a limitation of 50. This selection process allowed the identification of green spaces that were suitable for analysis.

Only 25 sites that had the most significant number of records and were over 100 observed difference from its adjacent sites were selected. The 25 sites were selected for two reasons. The first was for computing processing limitations. The second was that lower-ranking sites had larger gaps and temporal consistency in their datasets. Finally, previous studies in Taipei green spaces only an equal number of sites. Because of these factors, the process used eliminates the chance of sites with insufficient data for future analysis.

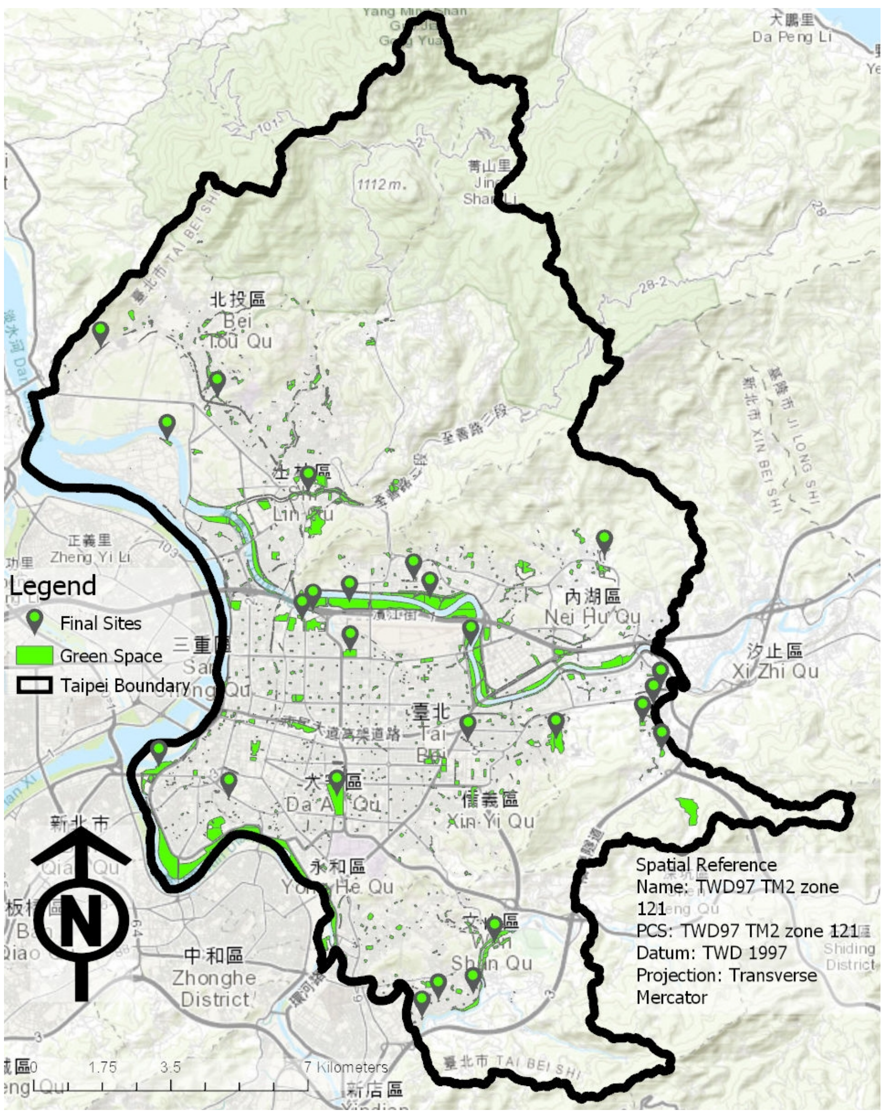


Figure 9: All 25 green spaces selected in Taipei

3.2.2. EBird Modifications

The avian dataset was modified using buffers for accounting for collection protocols. The standard protocol used in eBird is P25, where an observer collects data 30 meters from the point of collection. The new dataset was reviewed and edited before being used for further processes. Historical records and large numbers over an extensive period of time spent was omitted. This omission was completed using the SQL's Group By command and selecting one of the following columns found within eBird dataset: protocol, and time_expired. Protocols that were historical records that had over 100 observations in a single sitting were omitted. For time_exipred, data collection for a duration of +10 hours was omitted. This procedure will ensure quality in the dataset for future analysis.

A new dataset was created inside of PostgreSQL for ease of comparative analysis of the eBird dataset. This process was used as a junction table between eBird and BLI dataset. New values were used to connect the eBird dataset, and newly created junction table key, where values less than 700 were found only in BLI species checklist and values greater than 900 represent values only found in eBird's that were redundant nomenclature or unclear species names. The final database was connected with foreign keys to ensure the validating of values before analyzing with R (Figure 11).

3.3. Data Analysis

The final phase focused on analyzing datasets the modified and refined datasets. Using R, PostgreSQL and QGIS, richness and biodiversity were identified in the selected 25 sites. Three methods were to be used to determine the links between green spaces and avian diversity inside Taipei's green areas: rarefaction analysis, Shannon-Weiner biodiversity index and Jaccard's similarity index. Two libraries were used when working with R: Open Database Connectivity

(ODBC) and vegan libraries. The ODBI library was used for data exchange between R and PostgreSQL. All analyses used the vegan library imported into R. This library contained the R script to run the analyses. The rarefaction analysis allowed for a greater understanding of the optimal sample size. This is not needed with the Shannon's Index as its large variations of sample sizes to find comparative biodiversity of each site. The JSI allowed for a more in-depth understanding of each of the site to one another. Two variables were used with three different analysis: residency and temporal. Both these variables were explored were possible.

3.3.1 Rarefaction Analysis

The rarefaction analysis is a method to determine the richness of a specific area where one controls the amount of sample being randomly counted. This process finds the most common denominator of species richness through all the sites. The rarefaction also possesses a rarefaction curve. This curve is the average of all options of species accumulation in of a single site. When combined with multiple curves, it then can be used to find the most optimal sample size for collecting a multiple site's richness. When comparing multiple sites, the recorded counts may vary. Because of this disparity, some selected sites have the possibility of showing a higher richness because they have lower numbers to choose from, while sites with a larger recorded account could show a lower amount of richness. The rarefaction equation solves this discrepancy problem (Equation 1). The rarefaction takes a specified number of samples, the most common in all datasets, and compares the species richness at all locations. This allows for the most optimal pool size when comparing multiple sites that contain a larger difference in sample sizes.

$$E(S) = \sum_{i-1}^{S}\left(1 - \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}}\right]\right)$$

Eqaution 1:E(S) is rarefaction, S is the number of species, N_i the count of species i, and (N/n) is the binomial coefficient and N_i gives the probabilities that species i does not occur in a sample of size. (Gotelli and Colwell 2009).

PostgreSQL and R-Studio were both used in order to find a standardized richness with all sites. R-studio was connected to PostgreSQL using the ODBI library in R. This package allowed both applications to communicate and save information in a single database. Under the vegan library, the rarefy function allowed for the quick rarefaction analysis of eBird data of each of the sites (Oksanen, 2016). The output dataset was added to a newly created table for sites analysis.

3.3.2 Shannon–Weiner Biodiversity Index

The Shannon-Weiner biodiversity index determined the diversity based on abundance to support the notion concept of biodiversity in specific areas (Equation 2). This index, unlike the rarefaction analysis, assumes that all recorded species are the total species in the location. The pi in the equation can further be broken down to the amount of individuals of each species over the total number of individuals for a site. Because of this variable, the analysis makes use of the full dataset. This index also showed the evenness of the specific site, which was comprised of the species number and the abundance of the species. Since a bigger dataset could be at a disadvantage when using the rarefaction analysis as it will have sample size limitations, the Shannon-Weiner Index can be used to support the rarefaction analysis of richness in a specific area.

$$H' = -\sum_{i=1}^{S} p_i \ln p_i$$

Equation 2: The H’ is the Shannon-Weiner Index, p_i is the relative abundance of species i, S is the total number of species present and ln is the natural log (Oksanen 2016).

R-Studio and PostgreSQL were used for completion of the Shannon-Wiener Diversity analysis. For this analysis, the vegan package diversity function in R allowed the ability to compare multiple green spaces to find each of their avian diversity. The analysis results were stored in PostgreSQL producing a new dataset for the Shannon-Weiner biodiversity index.

3.3.3 Jaccard Similarity Index

The Jaccard similarity index was used to explore the similarity and diversity of a community (Equation 3). This index allowed for a clear understanding of multiple sites and the community composition of the avian species found within each site. The JSI equation focus on the relationship between intersections and unions of different groups or datasets. Because of this relationship, the X and Y repersents sets, where X∩Y is the intersect and XUY is the union of the sets. Understanding the intersections of multiple sets allows the ability to find composition similarities between different groups. This analysis was used with migratory avian species, resident species, as well as the combination of both avian groups to determine species if there were visible community guilds that have formed within the city of Taipei in specific green spaces.

$$J(X,Y) = \frac{|\mathbf{X} \cap \mathbf{Y}|}{|\mathbf{X} \cup \mathbf{Y}|}$$

Equation 3: J (X, Y) is the Jaccard similarity matrix. (Chao et al. 2006).

Data separated from PostgreSQL and running analysis on R was used to find JSI. The use of R's vegan library allowed for the calculation of JSI. The JSI was plotted using a custom command to make a dendrogram. This dendrogram determined the cluster of sites that have similar species composition. This plot was used to create a visual representation on a map using ArcGIS Pro.

3.3.4 Migration VS. Resident Avian Diversity

The ability to identify if these species are residents could give a greater understanding of the biodiversity of the individual site. Differentiation between migratory and resident species could affect the avian diversity in each of the green spaces as there could be dramatic population fluctuations during different periods. These fluctuations affect could both analysis methods and misrepresent the site. Because of the possibility of migratory avian species affecting the analysis, both migratory and resident avian species biodiversity analysis will be conducted.

PostgreSQL and SQL script were used to separate resident avian species from migratory and using the BLI created the dataset. This allowed for analysis in R-Studio between the two different groups. These were then transferred into R using the OBDC library where they were further analyzed (Figure 11).

3.3.5 Time Variable

Early analyses were used to determine the influence of diversity and richness on all the selected sites. However, significant temporal variations in the diversity count could be an indicator for inaccuracies in data. While biodiversity was determined in all the sites during all three years together, the study of the site data annually was taken. This allowed the study to explore temporal changes in selected areas. An overall analysis of recorded sightings completed of all the sites that contained enough recorded events. This showed the general avian biodiversity found through Taipei green spaces, yet this method does not give the full picture of these green spaces. Further examination of the dataset will be processed in R-Studio then saved in PostgreSQL to finally be explored in QGIS.

In the eBird dataset, only three years of data (2016, 2017 and 2018) were analyzed. Using the Shannon-Weiner diversity index and JSI methods described in Section 3.3.1 and 3.3.2, changes in biodiversity over time were analyzed. When working with the annual dataset, it must be noted that there were changes in the amount of observational data collected for each year. All results from the sites that were analyzed were saved in PostgreSQL and presented in QGIS.

Chapter 4 Results

A total of 112,871 events were recorded between 2016 and 2018. However, only 65,171 entries were selected due to multiple recordings having inaccurate species names. From the total of the selected entries, 237 avian species were found within the sites. Only 152 of the avian species were classified as migratory, and 95 were classified as resident species. Ten of the avian species were listed as threatened or endangered.

4.1. Rarefaction and Richness

Using the rarefaction analysis, there are two significant findings of the overall citizen-collected eBird dataset at the selected green space sites in Taipei. Rarefaction was able to determine the optimal point collection of all 25 of the sites (Figure 12). The optimal point is represented by the vertical line and was determined to be at 257 sample size for all sites. This means that if the sample size of each site is collected randomly at a designated observation amount, all sites have the most optimal levels for comparison when 257 observations were used. The second significant finding is the rarefaction curve. The curve shows that even with larger sample sizes there is a limitation to the richness. There is a clustering of richness that levels off around the sites that had only 50 to 100 species present. This supports the proposal that selected green spaces and species richness do not increase after this range.

Rarefaction eBird Dataset

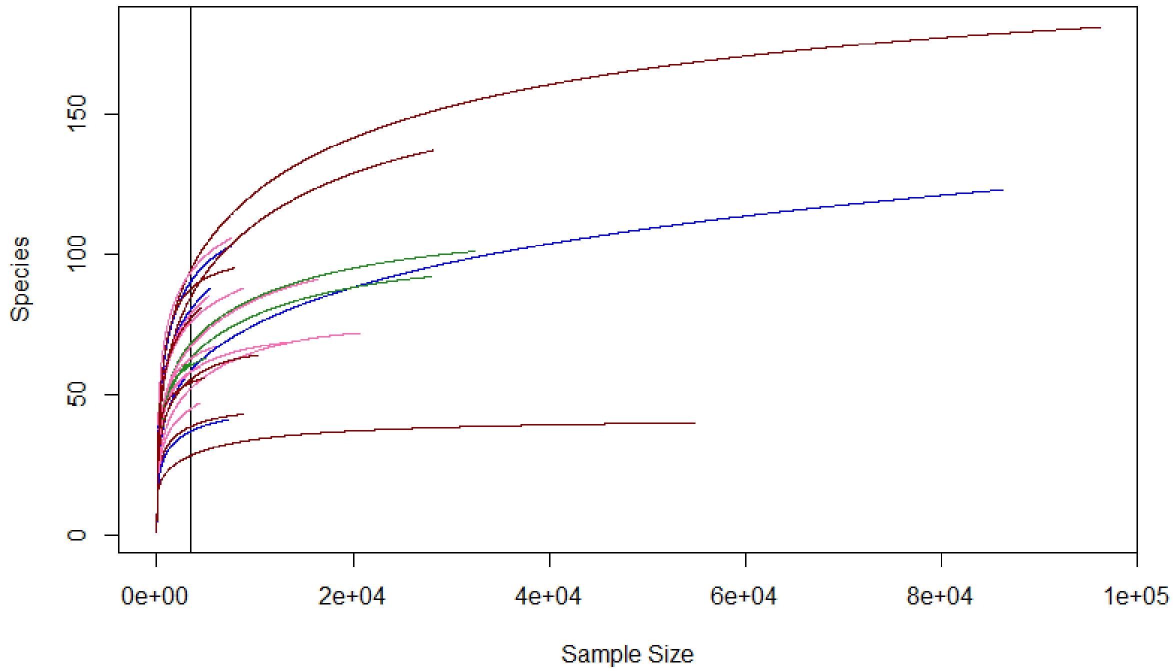


Figure 10: Refraction of eBird dataset from 2016 to 2018 using R

4.2. Dates Comparison

When using the Shannon-Weiner biodiversity index, the overall green spaces found within the study area showed a steady increase in biodiversity in all the sites (Figure 13). While there was an increase, there also seemed to be a limit of biodiversity. This limit is around 3.5 in the Shannon index. Between 2016 and 2018, there was a ten percent increase in the overall biodiversity throughout all sites, with the largest increase being from 2017 to 2018. The amount of biodiversity variation between sites decreased significantly in 2018. This could be an artifact of an increased amount of observations in these sites, which produced a more accurate biodiversity reading.

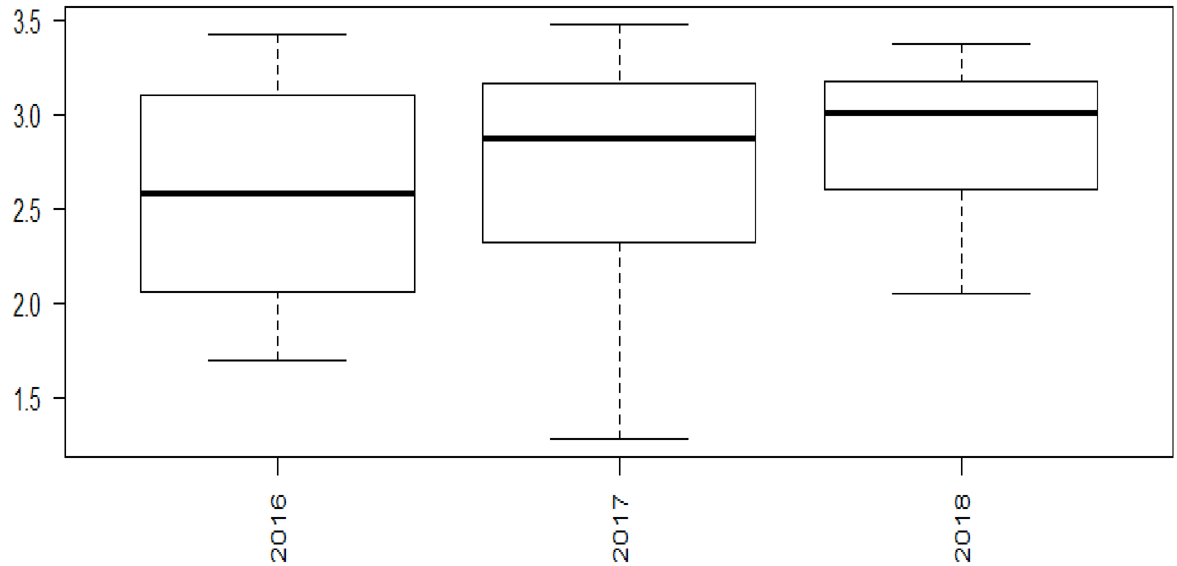


Figure 11: Averages of biodiversity of all sites between 2016 and 2018

Biodiversity had no significance by either location or sample size. The highest average diversity locations were found at sites with a large difference in the number of observations (Figure 14). Daan forest park, Shuagxi Riverside, and Mucha park had the highest numbers of observations; however, Daan was found to have a biodiversity index around 2.5. This is in contrast to the Shuagxi and Mucha park which averaged around 3. Unique green space attributes may be the largest factor affecting green space diversity. Sites with lower biodiversity tended to be more erratic in their yearly biodiversity. However, on average biodiversity increased in later dates of collection.

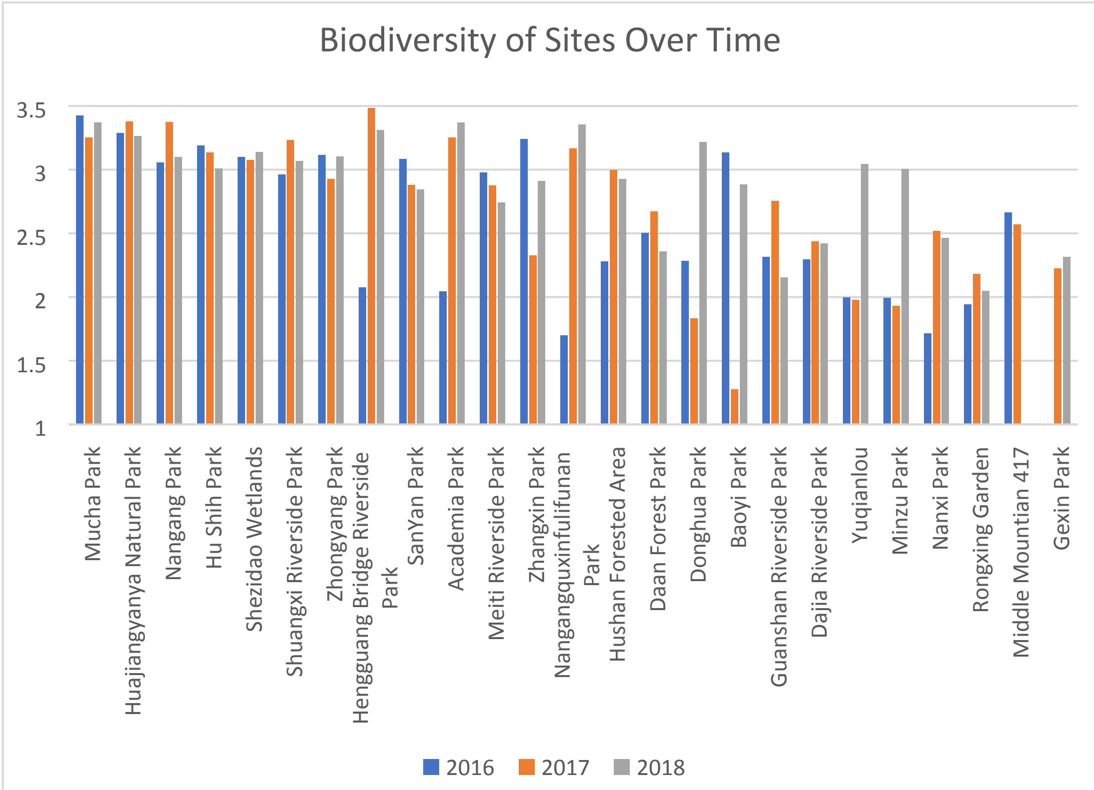


Figure 12: Biodiversity of each site over the years of 2016 to 2018

Patterns emerged over time. These patterns may not show an actual increase over the years, but instead stable biodiversity within the green space areas within Taipei. It is premature to say whether positive shifts in biodiversity are affecting many of these sites or if there are other variables relating to the site's attributes. While the most stable sites show the highest averages of biodiversity, this may be because of a temporal biased. Figure 13 showed that most sites in 2018 have higher biodiversity then previous years. It can be noted that some sites that had a lower biodiversity average had higher in biodiversity for the year 2018 then sites that had higher averages. Some of the mid-range sites found in the overall averages have stabilized with respect to their yearly biodiversity levels. Another result was that most of the dates in 2018 tended to be higher in the mid and lower biodiversity averages.

4.3. Biodiversity and Residency

The comparison between biodiversity and species richness is worth discussion (Figure14). Sites located nearer to the city center tended to have lower biodiversity compared to green spaces located closer to the edges of the administrative borders. This pattern implies that the different site attributes have an influence on their biodiversity. When comparing species richness to biodiversity, there is tends to be higher biodiversity when compared with higher richness. However, where species richness is below 50, this trend is weaker.

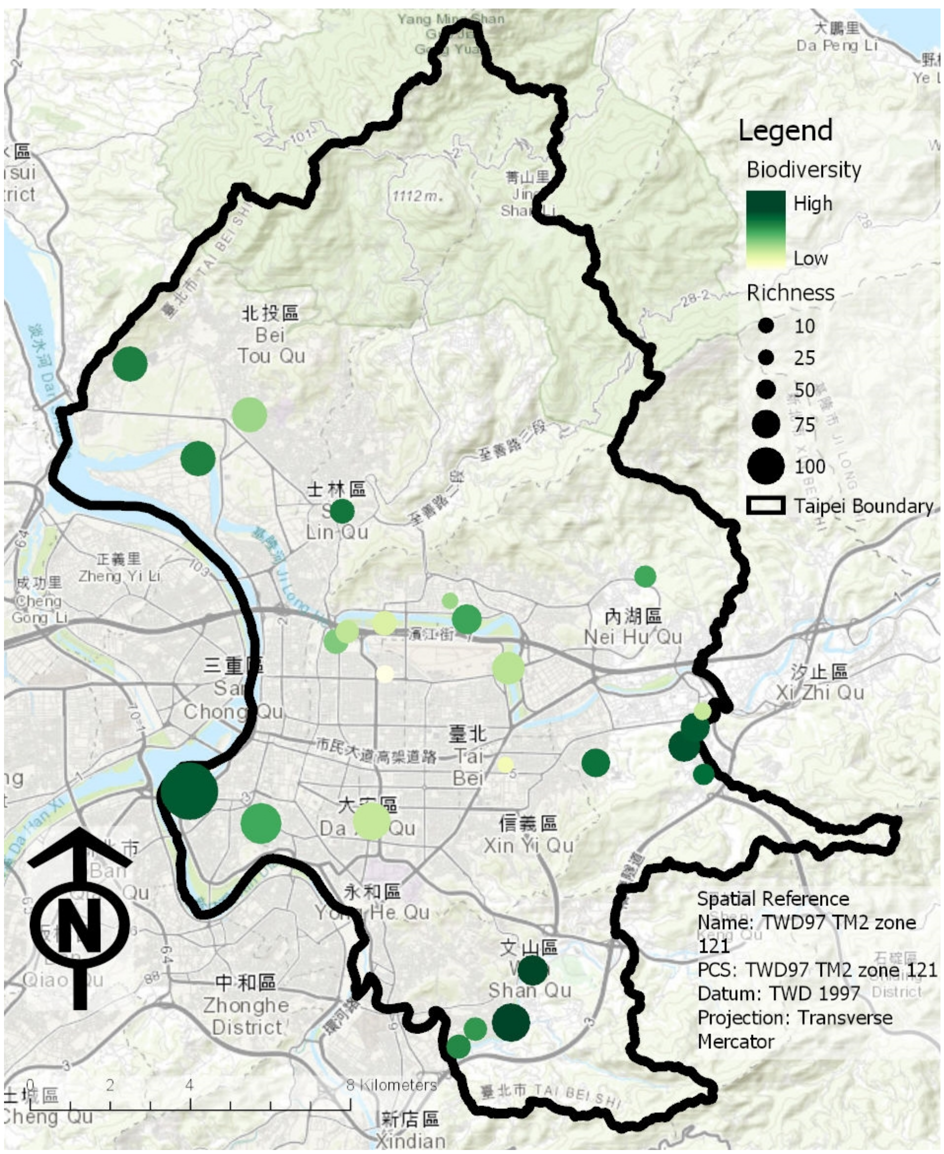


Figure 13: Collective Richness and Biodiversity

Biodiversity between residence and migratory avian species showed differences when viewed across the three years (Figure 15). The migratory species have higher biodiversity in most of the sites as compared with resident species. However, where there was a lack of water features, there tended to be an increase in resident species. This is different from migratory species, which tended to species decrease. This was seen with all separate dates noted from 2016 to 2018.

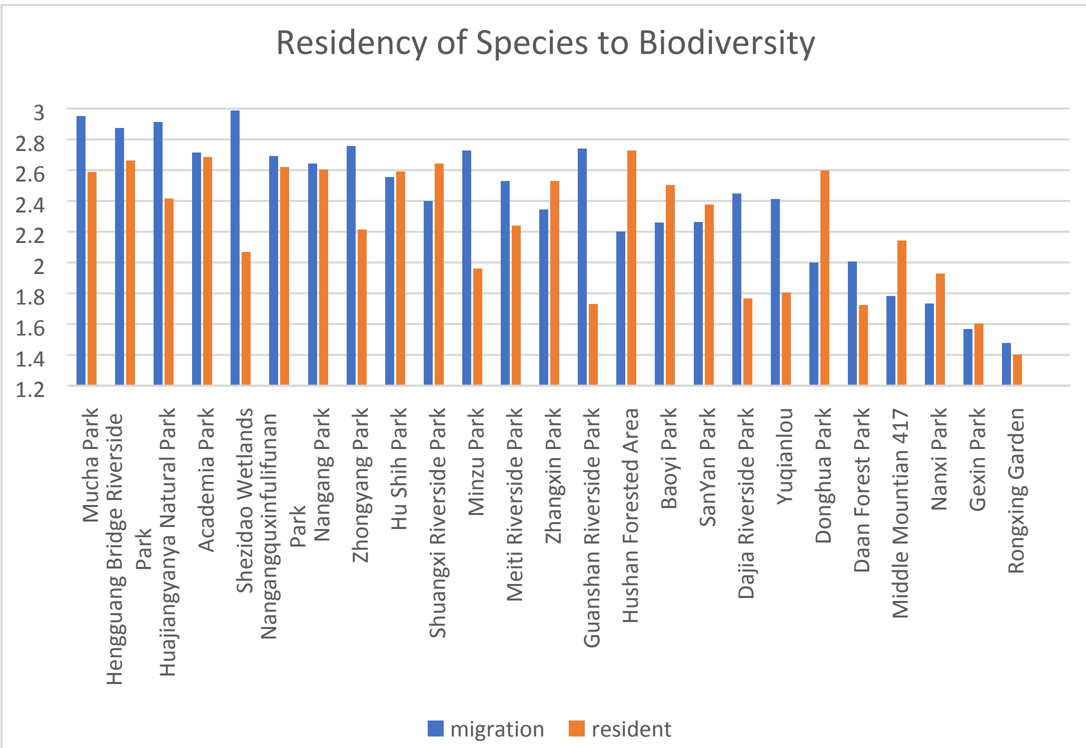


Figure 14: Migratory and resident species in all selected sites

4.4. Jaccard Clustering

There was a noticeable difference when comparing the migratory and residential avian species using the Jaccard clustering tool. The dendrograms showed variations between all three variables (Appendix D,E,F). However, the largest difference was with respect to migratory species as it was broken into four main groups. Compared to the other variables, the dendrogram only was divided into only three clusters. The species compositions were represented from blue to yellow depending on the similarities of each site composition (Figures 16, 17 18).

When exploring all the species using JSI, there were species composition similarities that become apparent (Figure 16). Sites closer to the city center had a similar species composition in comparison to sites located closer to the outskirts of the city. Both the south and the eastern sites have similar species composition. When richness and species composition was examined, there is a weak trend between sites species compositions and site richness.

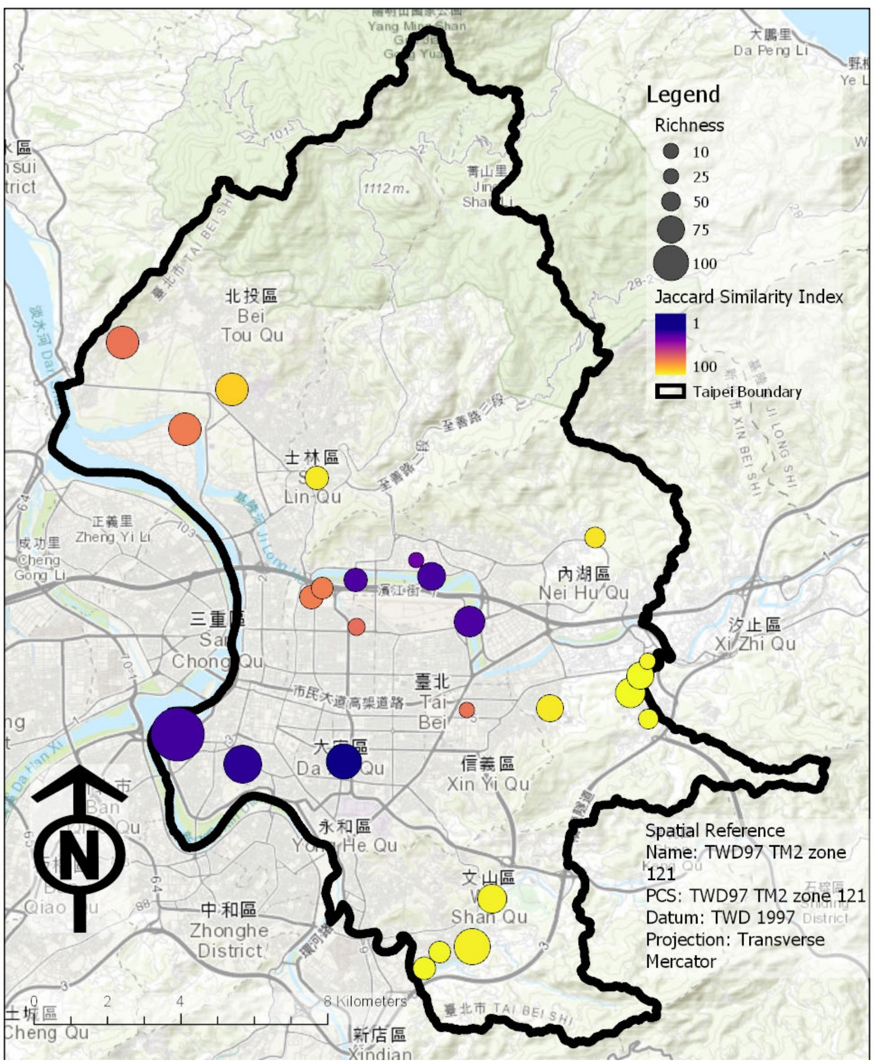


Figure 15: All species composition of each site. Each marker represents a site defined with size as richness and color as closeness of species composition of each site of all species between 2016 to 2018

The residential species between different sites had similar clustering with respect to the eBird dataset (Figure 17). However, individual sites within the city had closer species composition to species more often found outside of the city. Both the southern and eastern compositions were relatively similar when compared to the combined dataset.

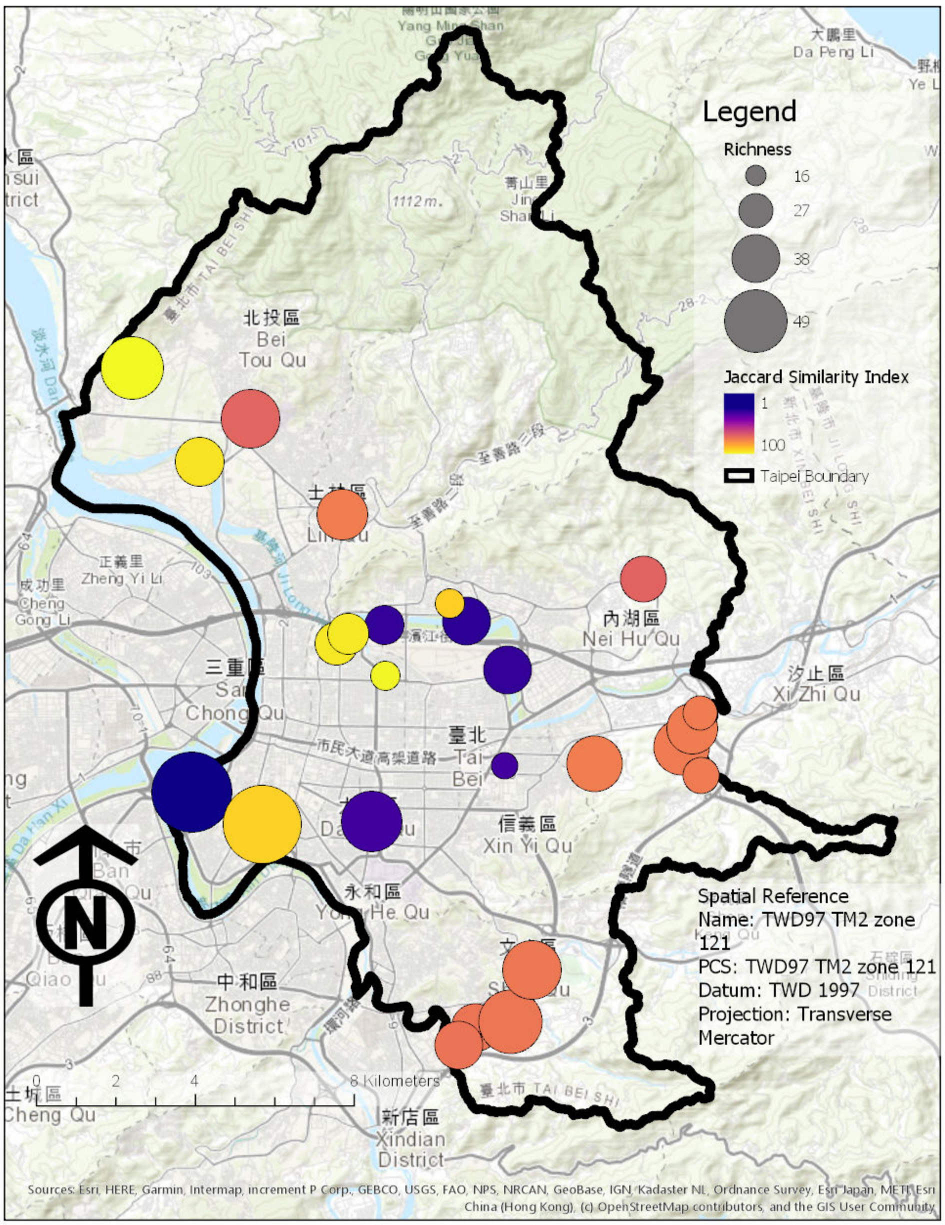


Figure 16: Resident species composition of each site. Each marker represents a site defined with size as richness and color as closeness of species composition of each site of all species between 2016 to 2018

Finally, the migratory species between different sites had similarities with residential species (Figure 18). All plotted maps within the city map had strong species composition similarities. Analogous to the resident species, there are clusters of similar migratory species within the city.

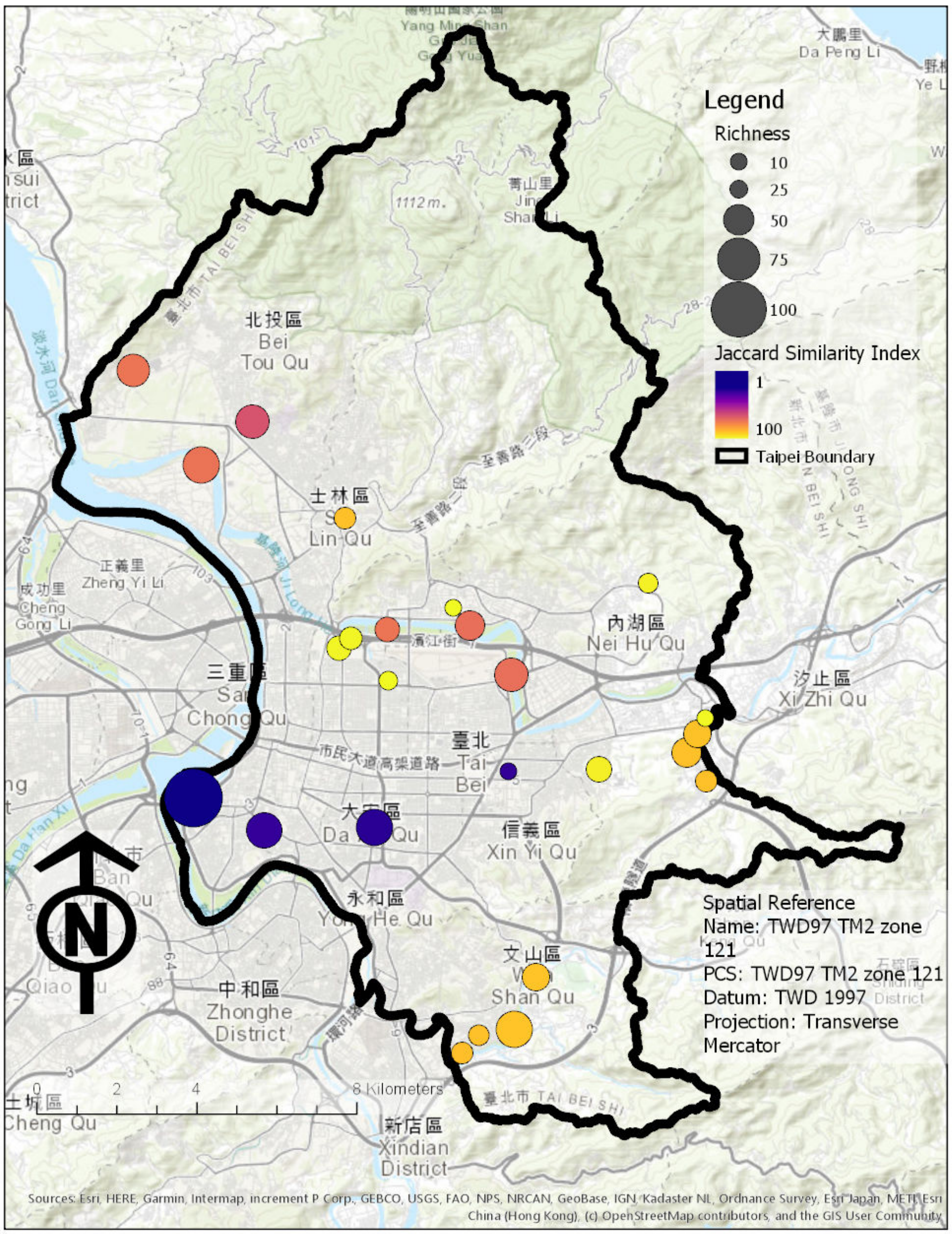


Figure 17: Migratory species composition of each site. Each marker represents a site defined with size as richness and color as closeness of species composition of each site of all species between 2016 to 2018

While all three plotted species compositions had similarities and differences, there was not enough information pointing to a strong trend. This outcome may be a result of the more considerable abundance of the resident avian species in comparison to the migratory species that were observed in the selected green spaces. While this appears to be a pattern among the selected sites, the sample size may be too small to generalize to the city of Taipei.

Chapter 5 Discussion

Monitoring the ecological health of green spaces within cities has consistently been a challenge for city planners and government officials. However, VGI datasets have allowed for a more accessible method for such monitoring compared to traditional means of data collection. It was determined with the use of eBird datasets that avian diversity in selected sites can be monitored with varying degrees of success.

5.1. Citizen-collected Datasets

The study indicated that citizen-collected data could monitor avian species in green spaces in the city of Taipei efficiently. Exploration of various factors inside the dataset has allowed for a more complete understanding of overall biodiversity and health of the city Taipei.

A distinct advantage in using citizen-collected datasets is the finding appropriate sampling size for a study. Compared to Shih's research, the citizen-collected dataset contained significantly more records where the study's sites overlapped (Shih 2010). This is seen when Shih's overall sample size for entire study was under 1000, while the citizen-collected from the eBird dataset had a minimum of 800 recorded events per site. These results also show a higher biodiversity average with respect to the citizen-collected datasets as compared to prior studies. When using the rarefaction analysis, it was determined that the optimal sample size was 257 for an individual site. This volume of data is due to the constant recordings through the three years the VGI dataset (Figure 19). This high volume of records would explain why the biodiversity averages for most of the sites were significantly higher in sites that of earlier studies. Because most sites were not able to achieve optimal sampling size using traditional collection methods, they may be prone to underestimate as site's avian biodiversity and community's composition.

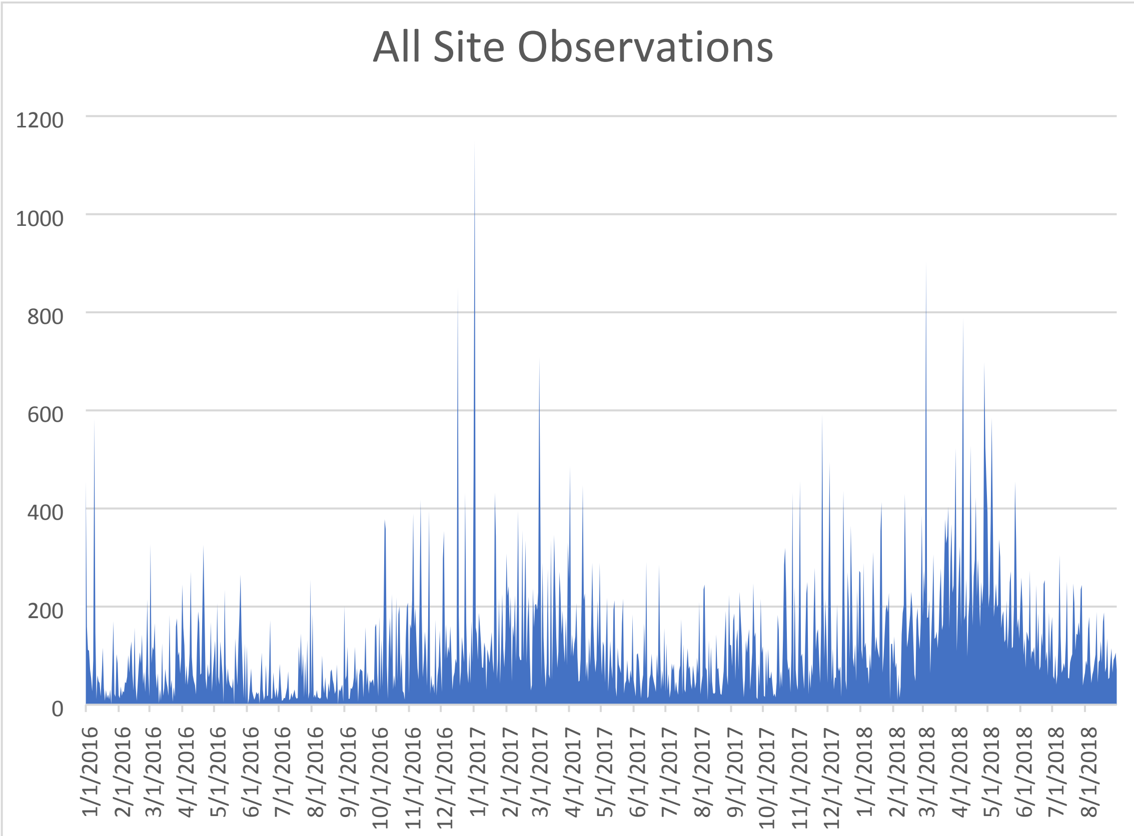


Figure 18: Observations recorded of eBird dataset between the dates of 2016 and 2018

It can be noted, while the observation count of the eBird dataset is high, there is a pattern. As Figure 16 showed, while biodiversity of all the selected sites is becoming higher over time, the observations have noticeable peaks and valleys. Most studies in Taipei where avian observations were used as an ecological health indicator were collected from July to November (Shih 2018; Shih 2010). If a VGI dataset used the same collection dates as previous studies, it would include some of the lowest recording frequencies. Because of these findings, the study supports the idea that VGI datasets allow for a higher chance of observing a selected green space's richness and biodiversity.

5.2. Different Variables

Analysis of the eBird dataset over three years allowed for a better understanding of green spaces within Taipei. Of all 25 selected green spaces, all but two of sites (Gexin and Middle

Mountain 417 Park) had readable levels of biodiversity for all three years. These sites reported low levels of average biodiversity and richness due to have significantly lower observer counts from a single individual yearly selection. However, the sites that had the overall highest average biodiversity tended to have the most stable year to year (Figure 13). Even with less stable years, 2017 and 2018 tended to be closer in biodiversity compared to earlier dates. Because of this trend, it can be supported that, VGI's eBird dataset showed levels of reliability when monitoring both richness and biodiversity of a selected green space as time progresses.

The analysis of resident and migratory species confirmed the there were some unique differences. While there was a higher number of migratory species, these species have a weak influence on the overall avian biodiversity of the selected sites. This might be due to many of the resident avian species being counted in higher number throughout the year while migratory species are only recorded when passing through on their yearly routes.

5.3. Species Composition

Species composition allowed for more in-depth exploration into both the migratory and resident avian species. There were no strong links between a site's species compositions and their richness. However, when visualizing the species compositions on the map (Figures 16, 17, 18), the sites located further from the urban centers and closer to the mountainous areas tended to have similar compositions. This clustering could be due to the attributes of the surrounding area. Geographical attributes such as elevation and proximity to structures could be affecting these compositions. This was most strongly seen in both the eastern and southern sites where the urban areas are less dense and closest to the mountainsides. Hillsides and heavily forest areas demonstrated to similar species clustering effect.

5.4. Limitations

The study was able to show trends using citizen-collected data, though there were notable limitations. The first limitation was the temporal dispersion of recorded events. Unlike traditional methods of avian data collection that determines a standard recording dataset, producing temporal evenness, citizen-collected datasets are collected on an observer's temporal bias. Because these datasets had peaks and valleys throughout the year and there was temporal unevenness through the varying sites (Figure 17). This is different compared to traditional studies that use the breeding seasons for all data entry. However, if used on this dataset it would have had a much lower Shannon's Index in most sites. This limitation could be addressed by either increasing the amount of years or having a standard time based on peaks and valleys. The second limitation is the multiple attributes that are inherent to individual selected sites. While the study did address the practical use of citizen-collected data, it did not examine the different factors that could be affecting their distribution. Attributes for both the locations and the species were limited in this study. Due to this lack of selected attributes, it would be premature to define as to the reason these trends are emerging.

Chapter 6 Conclusion

The objective of the thesis was to determine if citizen-collected datasets could be used as a reliable means of determining green space health in Taipei. Other studies have used avian surveys to conduct this process for the Taipei area in association with green spaces; none have used citizen-collected data. While the citizen-collected dataset allowed for high numbers of events, several of these were limited due to inaccuracy in species labels. This section summarizes the findings of the individual elements and elaborates on further exploration of future endeavors that could be taken.

6.1. Major Findings

The study confirmed that citizen-collected data can be used for determining a green space's ecological health using biodiversity and richness. The study determined a strong relationship between avian biodiversity and richness of green spaces in Taipei. However, there was only weak support between a site's species composition and richness. While there were limitations, the study was able to find the eBird's avian datasets can be used in monitoring multiple sites by a data-driven selection process throughout the municipality of Taipei.

Compared to traditional data collection, eBird has the benefit of simultaneously collecting data from various sites throughout a specific region. There were originally 458 green space parcels that correlated with the original criteria of this study. Only the 25 green spaces selected in Taipei were selected and included for this study. While there were a large number of green spaces within Taipei, many of these sites had too low of a count of observations to be accurately accessed. These irregularities can be caused by observational biases for specific locations. Even using sites with different sample sizes, there was a threshold to biodiversity and richness found in a specific green space. Identifying the biodiversity and richness thresholds found in these areas can be used in subsequent models as a baseline of overall environmental health.

The eBird dataset was shown to be reliable from 2016 to 2018 and shows promise of growth in the future. Many of the sites with the most entries tended to have steadily increasing biodiversity indexes for all three years. Reviewing how the eBird database has grown over the past ten years for Taipei, there is the expectation of the increase of usable data for more precise analysis in the future.

When comparing the migratory and resident species, migratory species tended to have higher species diversity and were found in more areas. This finding may because of a higher number of possible migratory species documented under the BLI checklist. The larger list could have shown greater bias for migratory species. When exploring community species composition, it was determined that many of the sites have a similar trend between the two specific categories. This was supported by the species found in the South and Southwestern green spaces.

6.2. Future Work

Although this study was able to examine the avian biodiversity of 25 sites found within urban areas of Taipei, there were many limitations, as discussed in Chapter 5, as to what could be ascertained from this individual study. With the ever-increasing amount of accessible data, some of these studies' limitation could be overcome.

With the more detailed data acquisition, an in-depth exploration of green space's ecological health could provide a richer understanding of the relationship of Taipei's green spaces and their biodiversity. All the sites within this study had a basic level of information designated to them (e.g. area, location, and type) much was not explored. Progressing from this study, multiple types of attributes will be essential to developing a more complete understanding of specific patterns. Collecting and examining information on types of vegetation, water features, and human factors will give a clearer picture of what could be affecting various species.

Having more data about different species allows for a better understanding of why different species inhabit specific green spaces. Knowing nesting and foraging habits can further understand the niches that these avian species occupy. This information would allow for a better understanding of a Taipei green space ecological health criteria as it would understand specific species limitations based on their different attributes. These new datasets could be updated inside PostgreSQL database table under avian_information. The current scope of the thesis did not require this information. If explored further, there could be more precise answers on why species compositions were presents and what is needed for them. This information could help create more suitable green spaces for multiple native species.

6.3. Overall Conclusions

This thesis was able to show a low-cost method to monitor the ecological health of 25 green spaces of Taipei. Using more accessible datasets, both city planners and citizens can have a greater understanding of their environment and use this information to develop urban policy and make planning decisions.

The objective of the study was to determine if the citizen collected dataset could be used to monitor the health of green spaces. A baseline of avian biodiversity was established during the exploration of various green space sites within the urban areas of the municipality of Taipei, Taiwan. Continuous data acquisition from the VGI datasets will allow the observation of trends in avian population and in the effectiveness of the green spaces in the urban areas over more extended periods of time. There is an understanding that in the future the trend of urbanization of the human population will require cities to plan for specific areas. While green spaces allow for a healthier citizenry and support of other organisms within metropolitan areas, they need to be monitored and better understood so that these limited areas are used to the utmost extent. Because of the ease and accessibility of VGI datasets, the potential they offer to monitor a city's valuable land resources will become ever increasingly important as more of the human population moves to the urban environment.

Reference

Aida, Nurul, Selvadurai Sasidhran, Norizah Kamarudin, Najjib Aziz, Chong Leong Puan, and Badrul Azhar. 2016. "Woody Trees, Green Space and Park Size Improve Avian Biodiversity in Urban Landscapes of Peninsular Malaysia." *Ecological Indicators* 69: 176–83.

Akbari, Hashem, and Dionysia Kolokotsa. 2016. "Three Decades of Urban Heat Islands and Mitigation Technologies Research." *Energy and Buildings* 133: 834–42.

Barbosa, Olga, Jamie A Tratalos, Paul R Armsworth, Richard G Davies, Richard A Fuller, Pat Johnson, and Kevin J Gaston. 2007. "Who Benefits from Access to Green Space? A Case Study from Sheffield, UK." *Landscape and Urban Planning* 83: 187–95.

"BirdLife." BirdLife. Accessed January 2, 2019. http://www.birdlife.org/ http://www.birdlife.org/.

Blair, Robert B. 1999. "Birds and Butterflies Along an Urban Gradient : Surrogate Taxa for Assessing Biodiversity ?" *Ecological Applications* 9 (1): 164–70.

Bouten, Willem, Edwin W. Baaij, Judy Shamoun-Baranes, and Kees C. J. Camphuysen. 2013. "A Flexible GPS Tracking System for Studying Bird Behaviour at Multiple Scales." *J Ornithol* 154: 571–80.

Chamberlain, D. E., S. Gough, H. Vaughan, J. A. Vickery, and G. F. Appleton. 2007. "Determinants of Bird Species Richness in Public Green Spaces." *Bird Study* 54 (1): 87–97.

Chandler, Mark, Linda See, Kyle Copas, Astrid M Z Bonde, Bernat Claramunt López, Finn Danielsen, Jan Kristoffer Legind, et al. 2017. "Contribution of Citizen Science towards International Biodiversity Monitoring."

Chang, Chi Ru, Ming Huang Li, and Shyh Dean Chang. 2007. "A Preliminary Study on the Local Cool-Island Intensity of Taipei City Parks." *Landscape and Urban Planning* 80 (4): 386–95.

Chao, Anne, Robin L. Chazdon, Robert K. Colwell, and Tsung Jen Shen. 2006. "Abundance-Based Similarity Indices and Their Estimation When There Are Unseen Species in Samples." *Biometrics* 62 (2): 361–71.

Chen, Xiao Ling, Hong Mei Zhao, Ping Xiang Li, and Zhi Yong Yin. 2006. "Remote Sensing Image-Based Analysis of the Relationship between Urban Heat Island and Land Use/Cover Changes." *Remote Sensing of Environment* 104 (2): 133–46.

Conrad, Cathy C, and Krista G Hilchey. 2011. "A Review of Citizen Science and Community-Based Environmental Monitoring: Issues and Opportunities." *Environmental Monitoring and Assessment* 176 (1–4): 273–91.

Estoque, Ronald C, Yuji Murayama, and Soe W Myint. 2017. "Effects of Landscape Composition and Pattern on Land Surface Temperature: An Urban Heat Island Study in the Megacities of Southeast Asia." *Science of the Total Environment* 577: 349–59.

Fink, Daniel, Wesley M Hochack, Benjamin Zuckerberg, and Steve T Kelling. 2011. "Modeling Species Distribution Dynamics with Spatiotemoral Exploratory Models: Discovering Patterns and Processes of Broad-Scale Avian Migrations." In *Procedia Environmental Sciences*, 7:50–55.

Fontana, Simone, Thomas Sattler, Fabio Bontadina, and Marco Moretti. 2011. "How to Manage the Urban Green to Improve Bird Diversity and Community Structure." *Landscape and Urban Planning* 101 (3): 278–85.

Fuller, Richard A, Katherine N Irvine, Patrick Devine-Wright, Philip H Warren, and Kevin J Gaston. 2007. "Psychological Benefits of Greenspace Increase with Biodiversity." *Biol. Lett* 3: 390–94.

Goddard, Mark A., Andrew J. Dougill, and Tim G. Benton. 2010. "Scaling up from Gardens: Biodiversity Conservation in Urban Environments." *Trends in Ecology and Evolution* 25 (2): 90–98.

Gotelli, Nicholas J, and Robert K Colwell. 2009. "Estimating Species Richness." In *Biological Diversity*, 39–54.

Guo, Guanhua, Zhifeng Wu, Rongbo Xiao, Yingbiao Chen, Xiaonan Liu, and Xiaoshi Zhang. 2015. "Impacts of Urban Biophysical Composition on Land Surface Temperature in Urban Heat Island Clusters." *Landscape and Urban Planning* 135: 1–10.

Huang, Kuo Ching, and Shih Laing Chan. 2014. "A Study on the Land-Cover Change Indicators of Taipei Metropolitan Areas." In *IOP Conference Series: Earth and Environmental Science*. Vol. 18.

Imai, Haruka, and Tohru Nakashizuka. 2010. "Environmental Factors Affecting the Composition and Diversity of Avian Community in Mid- to Late Breeding Season in Urban Parks and Green Spaces." *Landscape and Urban Planning* 96 (3): 183–94.

James, Fraces C, and Stephen Rathbun. 1981. "Rarefaction, Relative Abundance, and Diversity of Avian Communities." *The Auk* 98 (October): 785–800.

Kong, Fanhua, Haiwei Yin, Nobukazu Nakagoshi, and Yueguang Zong. 2010. "Urban Green Space Network Development for Biodiversity Conservation: Identification Based on Graph Theory and Gravity Modeling." *Landscape and Urban Planning* 95 (1–2): 16–27.

Li, Junxiang, Conghe Song, Lu Cao, Feige Zhu, Xianlei Meng, and Jianguo Wu. 2011. "Impacts of Landscape Structure on Surface Urban Heat Islands: A Case Study of Shanghai, China." *Remote Sensing of Environment* 115 (12): 3249–63.

Matsuba, Misako, Shota Nishijima, and Kazuhiro Katoh. 2016. "Effectiveness of Corridor Vegetation Depends on Urbanization Tolerance of Forest Birds in Central Tokyo, Japan." *Urban Forestry and Urban Greening* 18: 173–81.

McKinley, Duncan C., Abe J Miller-Rushing, Heidi L Ballard, Rick Bonney, Hutch Brown, Susan C Cook-Patton, Daniel M Evans, et al. 2017. "Citizen Science Can Improve Conservation Science, Natural Resource Management, and Environmental Protection." *Biological Conservation* 208: 15–28.

Melles, Stephanie, Susan M. Glenn, and Kathy Martin. 2003. "Urban Bird Diversity and Landscape Complexity: Species-Environment Associations Along a Multiscale Habitat Gradient." *Conservation Ecology*. Vol. 7. The Resilience Alliance.

"NLSC." National Land Surveying and Mapping Center. Last Modified 2004. https://www.nlsc.gov.tw/En.

Oksanen, Jari. 2016. "Vegan : Ecological Diversity." *Cran.R-Project*. https://doi.org/10.1029/2006JF000545.

Shiflett, Sheri A, Liyin L Liang, Steven M Crum, Gudina L Feyisa, Jun Wang, and G Darrel Jenerette. 2017. "Variation in the Urban Vegetation, Surface Temperature, Air Temperature Nexus." *Science of the Total Environment* 579: 495–505.

Shih, Wan-Yu. 2010. "Optimising Urban Green Networks in Taipei City: Linking Ecological and Social Functions in Urban Green Space Systems." University of Machester.

Shih, Wan Yu. 2018. "Bird Diversity of Greenspaces in the Densely Developed City Centre of Taipei." *Urban Ecosystems* 21 (2): 379–93.

Silvertown, Jonathan. 2009. "A New Dawn for Citizen Science." *Trends in Ecology and Evolution* 24 (9): 467–71..

Strohbach, Michael W, Susannah B Lerman, and Paige S Warren. 2013. "Are Small Greening Areas Enhancing Bird Diversity? Insights from Community-Driven Greening Projects in Boston." *Landscape and Urban Planning* 114: 69–79.

"Taipei DUD." 2009. Department of Urban Development. Department of Urban Development. July 26, 2009. https://english.udd.gov.taipei/Default.aspx.

Tajima, Kayo. 2003. "New Estimates of the Demand for Urban Green Space: Implications for Valuing the Environmental Benefits of Boston's Big Dig Project." *Journal of Urban Affairs* 25 (5): 641–55.

Tan, Jianguo, Youfei Zheng, Xu Tang, Changyi Guo, Liping Li, Guixiang Song, Xinrong Zhen, et al. 2010. "The Urban Heat Island and Its Impact on Heat Waves and Human Health in Shanghai." *International Journal of Biometeorology* 54 (1): 75–84.

"The World Bank." 2018. World Bank Group. 2018. http://www.worldbank.org/.

Tramer, Elliot J. 1969. "Bird Species Diversity: Components of Shannon's Formula." *Ecology* 50 (5): 927–29.

Tratalos, Jamie, Richard A Fuller, Philip H Warren, Richard G Davies, and Kevin J Gaston. 2007. "Urban Form, Biodiversity Potential and Ecosystem Services." *Landscape and Urban Planning* 83 (4): 308–17.

United Nations. 2011. *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*. United Nations. http://www.un.org/esa/population/publications/PopDistribUrbanization/PopulationDistributionUrbanization.pdf.

US EPA. 2018. "What Is Open Space/Green Space?" Accessed April 20, 2018.

https://www3.epa.gov/region1/eco/uep/openspace.html.

Weng, Qihao, Dengsheng Lu, and Jacquelyn Schubring. 2004. "Estimation of Land Surface Temperature–Vegetation Abundance Relationship for Urban Heat Island Studies." Accessed April 17, 2018. https://doi.org/10.1016/j.rse.2003.11.005.

Wiens, John, Robert Sutter, Mark Anderson, Jon Blanchard, Analie Barnett, Naikoa Aguilar-Amuchastegui, Chadwick Avery, and Stephen Laine. 2009. "Selecting and Conserving Lands for Biodiversity: The Role of Remote Sensing." *Remote Sensing of Environment* 113 (7): 1370–81.

Wolch, Jennifer R, Jason Byrne, and Joshua P Newell. 2014. "Urban Green Space, Public Health, and Environmental Justice: The Challenge of Making Cities 'Just Green Enough.'" *Landscape and Urban Planning* 125: 234–44.

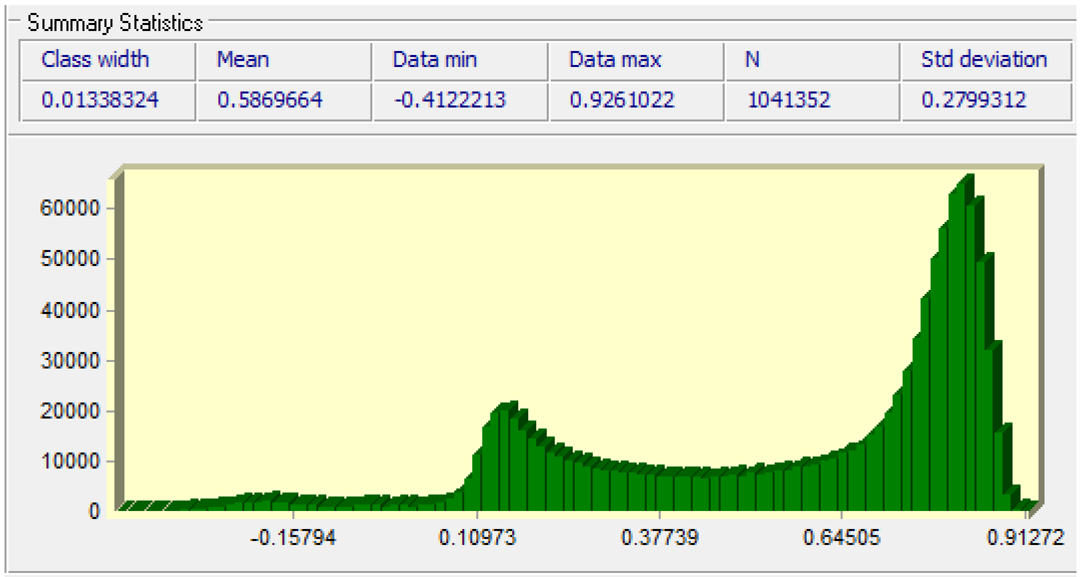
Appendix

```
SELECT lan.id, lan.geom,
        count(st_intersects(St_buffer(St_Transform(eb.geom,3826),30),
                    St_Buffer(St_Transform(lan.geom,3826),300)))  AS amount
    FROM landcover AS lan
    JOIN ebird AS eb
    ON St_Dwithin(lan.geom, eb.geom, 300)
    GROUP BY lan.id, lan.geom
    ORDER BY amount DESC
    LIMITS (N);
```

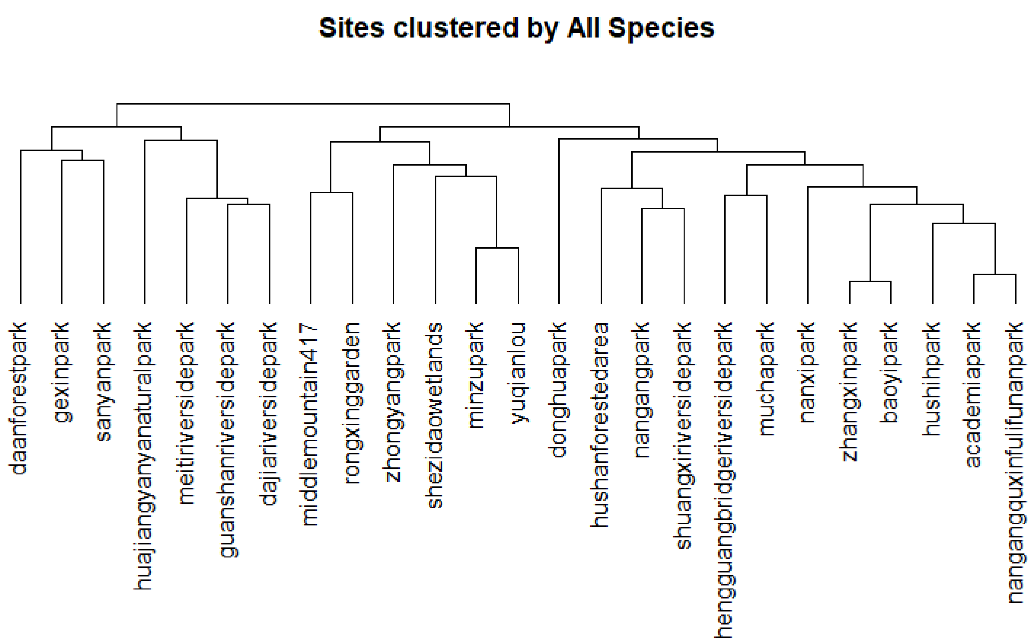
Appendix A: SQL query to determine suitable sites using modified landcover and eBird dataset. N represent then number of limitations. Both 300 and 50 were used for the value N.

```
WITH situation AS (SELECT * FROM ebird
 JOIN aviebird
 ON aviebird.aviannum = ebird.aviannum
 WHERE residency = 'M'
 AND ebird.aviannum < 900
 AND observat_1 BETWEEN '2016-01-01' AND '2018-12-31'),

            site1 AS (SELECT SUM(observatio), scientific
                FROM situation AS su
                JOIN sites AS s
                ON st_Dwithin(s.geom, su.geom, 300)
                WHERE s.id = 1
                GROUP BY su.scientific),


            site(N) AS (SELECT SUM(observatio), scientific
                FROM situation AS su
                JOIN sites AS s
                ON st_Dwithin(s.geom, su.geom, 300)
                WHERE s.id = (N)
                GROUP BY su.scientific)

SELECT DISTINCT situation.scientific, COALESCE(site1.sum, 0) AS site1,
                        COALESCE(site2.sum, 0) AS site(N),
        From situation
        LEFT JOIN site1 ON (situation.scientific = site1.scientific)
        LEFT JOIN site2 ON (situation.scientific = site(N).scientific)
            ORDER BY situation.scientific;
```

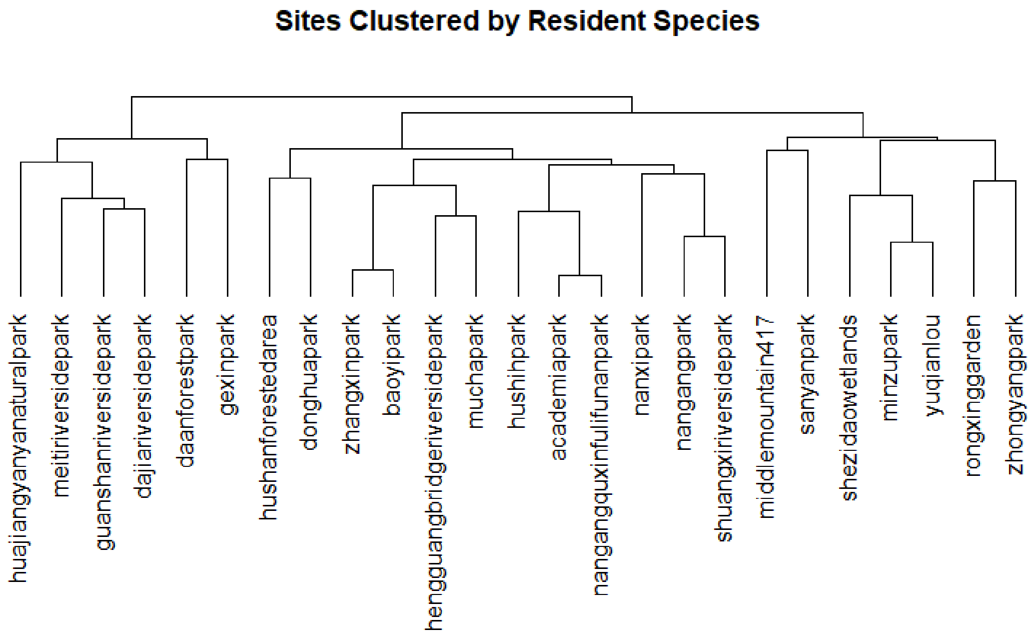
Appendix B: SQL script for transferring variables into R from PostgreSQL



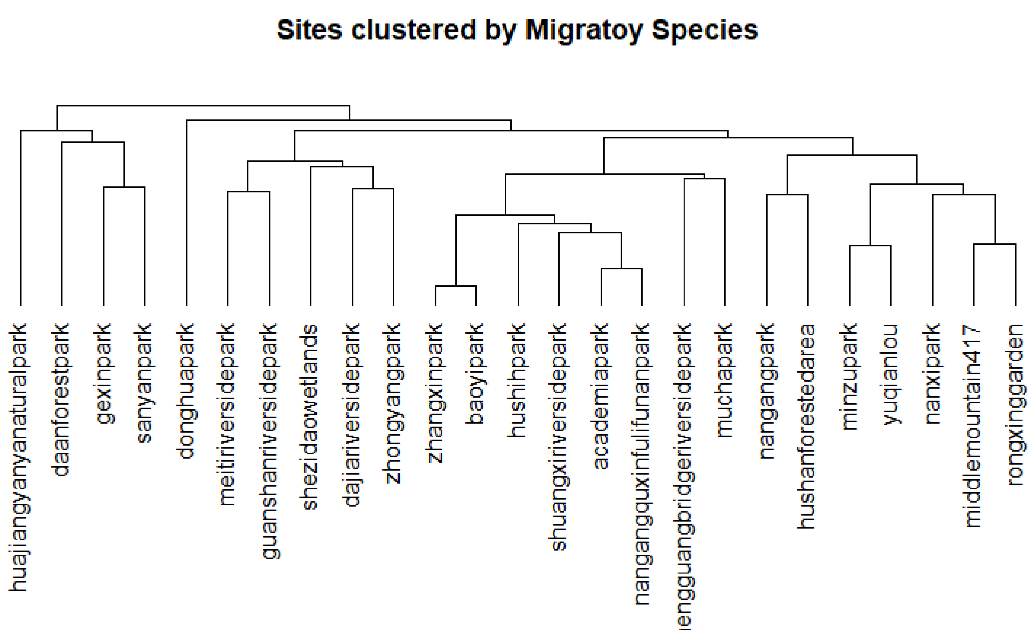
Appendix C: Histogram of NDVI dataset



Appendix D: Dendrogram JSI of all species from all selected site



Appendix E: Dendrogram JSI of resident species from all selected site



Appendix F: Dendrogram JSI of migratory species from all selected site