Database Design, Implementation, and Application: Bird Sightings in Dane County, Wisconsin

Barbara Heindl (M.S. Cartography & GIS) & Ellie Milligan (M.S. Cartography & GIS)
Geography 574, Dr. Qunying Huang
December 17, 2018

Problem Statement

It is well known that many wildlife species are intricately connected to the habitat that they exist in (Chase). Information about the preferred habitats of individual species is crucial in the development of management plans for those species (Crampton). As the landscape changes due to climate change and habitat conversion/ urbanization many species will be extirpated from their original home ranges or find themselves isolated from other populations (Atkinson) . Translocations of species from traditional home ranges where habitat has become sub par to areas where appropriate habitat is present but under utilized by that species has become a common technique for wildlife managers to bolster and protect species from going extinct (Conant). Another technique is to restore degraded habitat into more suitable habitat through restoration techniques and management. In order to use any of these techniques for a particular target wildlife species it is important to understand the basic needs of that species including current and historical range to understand how their distribution is changing and also what landscape and habitat features are requirements for a target species. Restoration and translocation efforts require a huge amount of monetary and logistical resources and without the proper background work done to assure the success of an effort a projects failure might be a huge loss in those resources and also to the species the project is attempting to help (Crampton) . As such, it is absolutely critical to first understand how a species interacts with its habitat. This can be done through niche modeling and through extensive research of where and when the species has been detected over time. We set out to create a database that would incorporate bird detections, several landscape and habitat features at those detections and weather/climate data near those detections sites. Searches and queries done through the database can be set up to search for a particular species to look at historical phenology variations or to explore landscape qualities consistently present when species are detected. Queries can also be done based on landscape features such as particular water features, land classes or edge habitat to look at the species associated in those areas. The database can also be used to look at the common species in an area based on season (Spring, Fall, Winter, Summer) or by temperature, precipitation and snow cover most common during detections. Due to time and size constraints our database will focus on sightings and climate data for Dane County, Wisconsin but could be expanded easily to include other states and countries.

Research Questions

1. Given a species, has the spring arrival date changed over time?
2. Do temperature, precipitation, or other climate-related factors appear to affect the first yearly detection of a species?
3. What is the detection density of a particular species in Dane County? In Madison?
4. What landscape variables are consistent with the detection of a certain species?

Data Sources

eBird: eBird is an online repository of bird sighting data collected via citizen science around the world. It aids in the collection of information related to species abundance, habitats, and other trends, with quality control overseen by Cornell University. In addition to species sightings eBird compiles a number of landscape and habitat measures at reported survey areas using landsat imagery. Both survey (bird detection) and landscape (covariate) data from eBird are available to the public by request. [eBird. 2017. eBird: An online database of bird distribution and abundance [web application]. eBird, Cornell Lab of Ornithology, Ithaca, New York. Available: http://www.ebird.org. (Accessed: December, 3 2018).]

Midwest Regional Climate Center: The MRCC is a program consisting of cooperations between the University of Illinois at Urbana-Champaign and several federal government agencies, including NOAA, the National Weather Service, and the USGS. They work to provide detailed historical climate data and data summaries based on reports from weather stations throughout the Midwest. [https://mrcc.illinois.edu/ (Accessed: December 4, 2018)]

Database Design

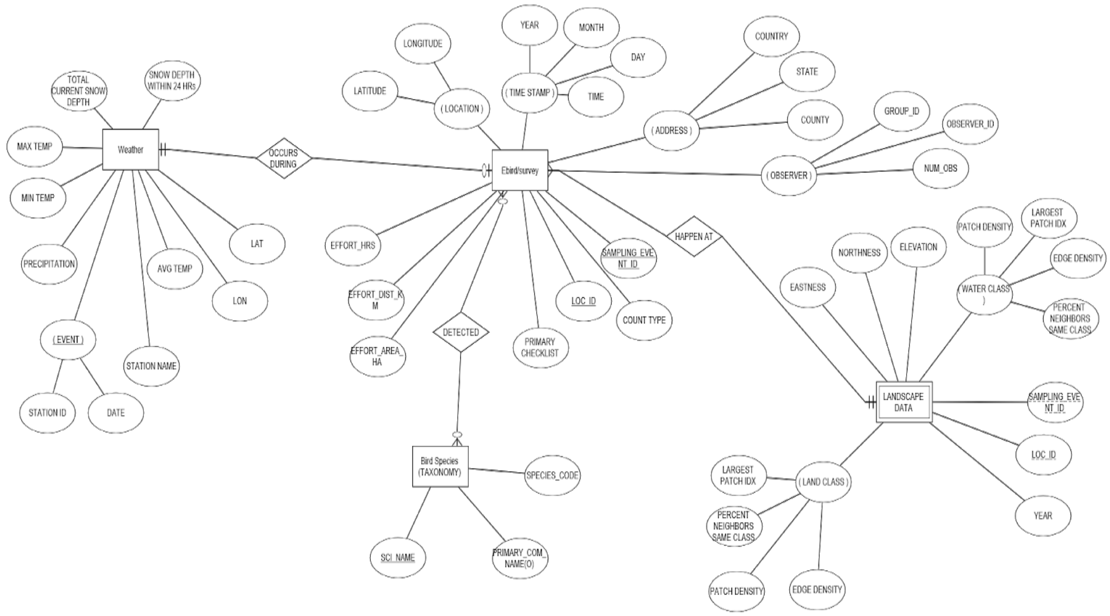


Figure 1: ER Diagram

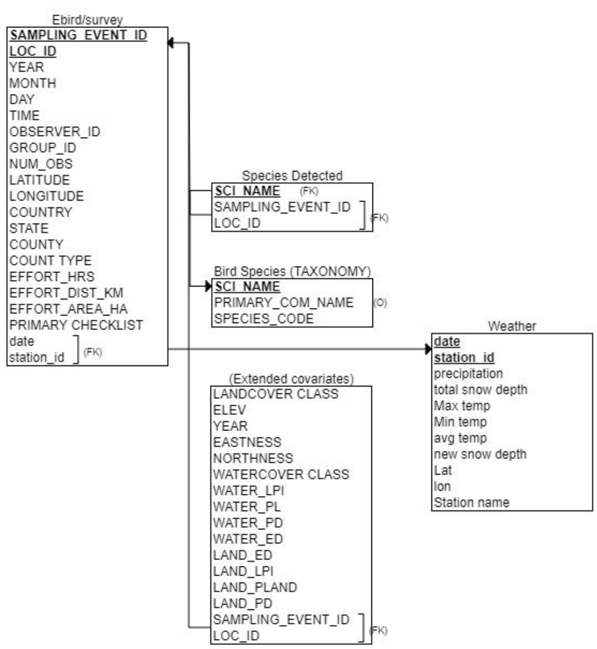


Figure 2: Relational diagram

Database Implementation

We used pgAdmin 4, an open source interface designed for use with PostgreSQL object-oriented databases to create tables and perform queries based on our original objectives and data sources. We then imported query results into QGIS (2.14) to map results for easier dissemination of results.

Table Descriptions

eBird Taxonomy: This table includes columns for the scientific name, the primary common name, and the species ID of each species listed in eBird. The species scientific name acts as the unique primary key.

Variable Name	Description	Data type
Scientific Name	Genus_species. Primary Key.	Varchar(50)
Common Name	Full common name.	Varchar(50)
Species Code	Code supplied by ebird.	Varchar(50)

eBird Survey: The eBird survey is the data filled out by the bird sighters and is verified by Cornell University. It includes sampling ID and location ID, which act as a composite key, as well as the latitude and longitude of the sighting, species scientific names, number of detected birds of each species, and information about the effort put into the survey by the observer. Since this data was imported as a CSV, we created the geometry based on the given latitude and longitude.

Variable Name	Description	Data type
Sampling Event ID	Unique identifier for each data sample / checklist. Primary Key	Varchar(15)
Location ID	Unique identifier for each observation location. Primary Key	Varchar(12)
Latitude	Decimal latitude. Location is tied to starting	Numeric

	position of traveling counts. Datum = WGS84.	
Longitude	Decimal longitude. Datum = WGS84.	Numeric
Year	Year of observation	Integer
Month	01- 12	Integer
Day	Day of the year 1- 366	Integer
Time	Time Observation began (0:00 - 24:00) Fractional hours. Based on time zone of observation.	Numeric
Country	Full name of political unit.	Varchar(30)
State or Province	Full name of administrative region.	Varchar(14)
County	Name of county or sub admin region.	Varchar(12)
Count type	Code given based on survey type (see appendix).	Varchar (3)
Effort Hours	Duration of observation in hours.	Numeric
Effort Distance	Distance traveled from observation start in kilometers 0 = stationary observation.	Numeric
Effort Area	Size of area covered during observation 0 = non area survey.	Numeric
Observer ID	Unique ID of person who submitted data	Varchar(9)
Number of Observers	Number of observers on survey	Integer
Group ID	Unique ID indicating multiple ebird users on same survey	Varchar (8)
Primary checklist	Boolean value used for ebird purposes	varchar(1)
Species observered	Numeric value of number of an individual species for all extant WI species. 0 indicates the species was not seen, blank indicates data for the species was unavailable.	Numeric
Geometry	Populated in pgadmin 4	Geom

Table 1. Sampling event covariates provided by eBird. (Fink)

eBird Covariates: The covariate table includes a wide variety of landscape information associated with the locations at which birds are sighted by eBird observers. This includes elevation, slope, extant land cover and water classes, edge density, patch density, and surrounding neighborhood metrics information among other variables. Measures for these variables are provided by eBird and derived using a variety of data sources including UMD land and QA water classes and NASA MODIS land cover imagery and was determined based on imagery taken during the year of the survey (Fink).

Variable Name	Description	Data type
Sampling event ID	Unique identifier for each data sample / checklist. Foreign Key which matches the sampling event table.	Varchar(15)
Location ID	Unique identifier for each observation location. Foreign Key which matches the sampling event table.	Varchar(12)
Eastness	A measure of slope and aspect.	Numeric
Elevation	In meters.	Numeric
Water Class Edge density	Edge density for each water cover type in the survey area.	Numeric
Water Class Largest Patch Index	Percentage of survey neighborhood (500m pixels) which is comprised of the largest patch of each water cover type.	Numeric
Water Class Patch Density	Number of Patches of each water cover type per 100 hectares in surrounding neighborhood.	Numeric
Water Class Percent	Percent of surrounding neighborhood that is each water cover type.	Numeric
Northness	A measure of slope and aspect.	Numeric
Land Class Edge Density	Edge density for each land class type.	Numeric
Land Class Largest Patch Index	Percentage of survey neighborhood (500m pixels) which is comprised of the largest patch of each land class type.	Numeric

Land Class Patch Density	Number of Patches of each land class type per 100 hectares in surrounding neighborhood.	Numeric
Land Class Percent	Percent of surrounding neighborhood that is each land class type.	Numeric
Land Cover Type	Class code for survey area (see appendix 2)	Integer
Watercover	Water cover code for survey area (see appendix 2)	Integer
Year	Year of observation	Integer

Table 2. Extended covariate/ landscape data provided by eBird (Fink). Full descriptions can be found in Appendix 3.

Dane County Climate: This information was sourced from the Midwest Regional Climate Center. We only collected climate data from Dane County due to the fact that all the data was stored by individual weather station. We collected the station data from the University of Wisconsin Arboretum, Dane County Regional Airport, Charmany Farm, Mazomanie, Middleton, Mt. Horeb and Stoughton, as all but one of these stations had relatively complete data (in terms of collecting information on both temperature and precipitation) all the way back to 2002, which is the first year for which we have data from eBird.

Variable Name	Description	Data Type
Station Name	Full name.	Varchar (50)
Station ID	Unique identifier. Composite Key.	Varchar(15
Latitude	Decimal latitude. Datum = WGS84.	Numeric
Longitude	Decimal longitude. Datum = WGS84.	Numeric
Date	YYYY-MM-DD Composite Key.	Date

Precipitation	Precipitation within in the previous 24 hr in inches. T = trace	Varchar (10)
Snow	Snow accumulation within the previous 24 hr in inches. T = trace	Varchar(10)
Total Snow Depth	Total snow accumulation present in inches.	Varchar(10)
Minimum Temperature	In the past 24 hr. Measured in Fahrenheit.	Varchar(10)
Maximum Temperature	In the past 24 hr. Measured in Fahrenheit.	Varchar(10)
Mean Temperature	In the past 24 hr. Measured in Fahrenheit.	Varchar(10)
Geometry	Populated in pgadmin 4	Geom

Table 3. Weather/ Climate data as obtained by MRCC.

Case Studies

Query 1: On what days was a particular species sighted between 2004 and 2016?

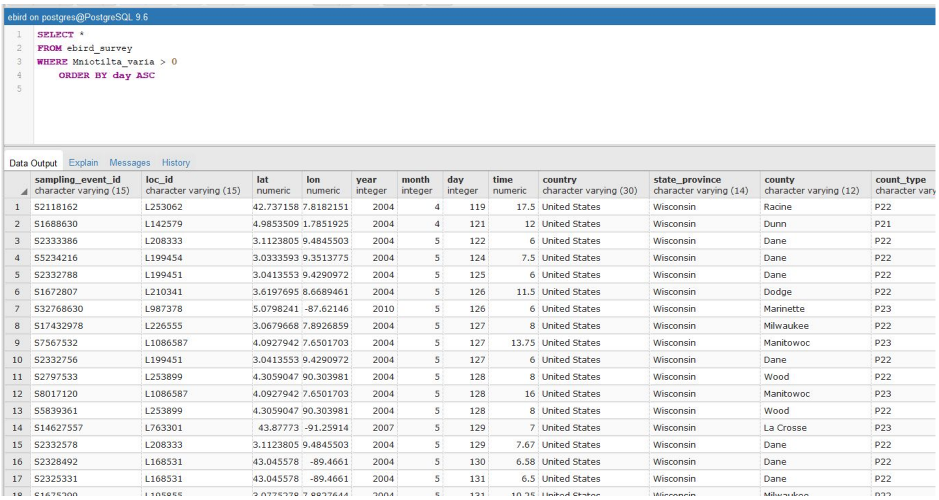
This query is useful for seeing the change in detection dates for a given species over the years, as it lists each individual day where a bird sighting was registered with eBird. eBird lists days in terms of the Julian Date of the year (i.e. January 1st is equal to 1, February 2nd is equal to 33, etc.). For example, we can see the the black-and-white warbler, a migratory song bird, was detected much earlier in the year in 2004 than it was in 2007. This gives us an indication that it may be worth looking into the climate data related to these dates to see if there was any factor that may have influenced this change in detection date.

SELECT *

FROM ebird_survey

WHERE Mniotilta_varia > 0

ORDER BY day ASC



	sampling_event_id character varying (15)	loc_id character varying (15)	lat numeric	lon numeric	year integer	month integer	day integer	time numeric	country character varying (30)	state_province character varying (14)	county character varying (12)	count_type character vary
1	S2118162	L253062	42.737158	7.8182151	2004	4	119	17.5	United States	Wisconsin	Racine	P22
2	S1688630	L142579	4.9853509	1.7851925	2004	4	121	12	United States	Wisconsin	Dunn	P21
3	S2333386	L208333	3.1123805	9.4845503	2004	5	122	6	United States	Wisconsin	Dane	P22
4	S5234216	L199454	3.0333593	9.3513775	2004	5	124	7.5	United States	Wisconsin	Dane	P22
5	S2332788	L199451	3.0413553	9.4290972	2004	5	125	6	United States	Wisconsin	Dane	P22
6	S1672807	L210341	3.6197695	8.6689461	2004	5	126	11.5	United States	Wisconsin	Dodge	P22
7	S32768630	L987378	5.0798241	-87.62146	2010	5	126	6	United States	Wisconsin	Marinette	P23
8	S17432978	L226555	3.0679668	7.8926859	2004	5	127	8	United States	Wisconsin	Milwaukee	P22
9	S7567532	L1086587	4.0927942	7.6501703	2004	5	127	13.75	United States	Wisconsin	Manitowoc	P23
10	S2332756	L199451	3.0413553	9.4290972	2004	5	127	6	United States	Wisconsin	Dane	P22
11	S2797533	L253899	4.3059047	90.303981	2004	5	128	8	United States	Wisconsin	Wood	P22
12	S8017120	L1086587	4.0927942	7.6501703	2004	5	128	16	United States	Wisconsin	Manitowoc	P23
13	S5839361	L253899	4.3059047	90.303981	2004	5	128	8	United States	Wisconsin	Wood	P22
14	S14627557	L763301	43.87773	-91.25914	2007	5	129	7	United States	Wisconsin	La Crosse	P23
15	S2332578	L208333	3.1123805	9.4845503	2004	5	129	7.67	United States	Wisconsin	Dane	P22
16	S2328492	L168531	43.045578	-89.4661	2004	5	130	6.58	United States	Wisconsin	Dane	P22
17	S2325331	L168531	43.045578	-89.4661	2004	5	131	6.5	United States	Wisconsin	Dane	P22

Query 2: What covariates are associated with sightings of a particular bird species?

This query links the eBird survey table with the covariate data table, allowing the user to see what landscape information was associated with a particular bird sighting on a particular day. These two tables are inherently related but were separated for ease of data management, so this operation brings them back together for a comprehensive understanding of a detection. This particular query allows us to see that Wood Ducks (*Aix sponsa*) a unique water bird for the fact that it perches in trees, has detections consistently with areas that contain landclass types that are wooded and associated with water classes that include lake shore.

SELECT *

FROM ebird_survey as ebird, covariates

WHERE ebird.aix_sponsa > 0 and ebird.loc_id=covariates.loc_id AND
ebird.county=`Dane`

ebird on postgres@PostgreSQL 9.6

```sql
SELECT *
FROM ebird_survey as ebird, covariates
WHERE ebird.aix_sponsa > 0 and ebird.loc_id=covariates.loc_id and ebird.county='Dane'
```

Data Output | Explain | Messages | History

	sampling_event_id character varying (15)	loc_id character varying (15)	lat numeric	lon numeric	year integer	month integer	day integer	time numeric	country character varying (30)	state_province character varying (14)	county character varying (12)	count_type character varying (3)	effort_hrs numeric	effort_ numeric
1	S1668414	L208954	3.0864413	9.4251679	2004	4	119	8.5	United States	Wisconsin	Dane	P22	3	
2	S1668228	L208954	3.0864413	9.4251679	2004	4	94	8	United States	Wisconsin	Dane	P22	1.5	
3	S1668414	L208954	3.0864413	9.4251679	2004	4	119	8.5	United States	Wisconsin	Dane	P22	3	
4	S1668228	L208954	3.0864413	9.4251679	2004	4	94	8	United States	Wisconsin	Dane	P22	1.5	
5	S2877785	L208333	43.11238	-89.48455	2007	4	96	7.42	United States	Wisconsin	Dane	P22	0.5	
6	S2871554	L208333	43.11238	-89.48455	2007	4	92	16.25	United States	Wisconsin	Dane	P22	3.417	
7	S2871234	L208333	43.11238	-89.48455	2007	4	92	16.5	United States	Wisconsin	Dane	P21	1.75	
8	S2204578	L208333	3.1123805	9.4845503	2004	5	138	8.75	United States	Wisconsin	Dane	P22	1.25	
9	S1659280	L208333	3.1123805	9.4845503	2004	4	101	9.5	United States	Wisconsin	Dane	P22	3	
10	S1659126	L208333	3.1123805	9.4845503	2004	4	101	9	United States	Wisconsin	Dane	P22	3	
11	S2877785	L208333	43.11238	-89.48455	2007	4	96	7.42	United States	Wisconsin	Dane	P22	0.5	
12	S2871554	L208333	43.11238	-89.48455	2007	4	92	16.25	United States	Wisconsin	Dane	P22	3.417	
13	S2871234	L208333	43.11238	-89.48455	2007	4	92	16.5	United States	Wisconsin	Dane	P21	1.75	
14	S2204578	L208333	3.1123805	9.4845503	2004	5	138	8.75	United States	Wisconsin	Dane	P22	1.25	
15	S1659280	L208333	3.1123805	9.4845503	2004	4	101	9.5	United States	Wisconsin	Dane	P22	3	
16	S1659126	L208333	3.1123805	9.4845503	2004	4	101	9	United States	Wisconsin	Dane	P22	3	
17	S2877785	L208333	43.11238	-89.48455	2007	4	96	7.42	United States	Wisconsin	Dane	P22	0.5	
18	S2871554	L208333	43.11238	-89.48455	2007	4	92	16.25	United States	Wisconsin	Dane	P22	3.417	
19	S2871234	L208333	43.11238	-89.48455	2007	4	92	16.5	United States	Wisconsin	Dane	P21	1.75	
20	S2204578	L208333	3.1123805	9.4845503	2004	5	138	8.75	United States	Wisconsin	Dane	P22	1.25	
21	S1659280	L208333	3.1123805	9.4845503	2004	4	101	9.5	United States	Wisconsin	Dane	P22	3	
22	S1659126	L208333	3.1123805	9.4845503	2004	4	101	9	United States	Wisconsin	Dane	P22	3	

Query 3: Which bird species were detected within a certain distance of a weather station?

This query allows us to see how climate data is associated with bird detections.

SELECT ebird.Mniotilta_varia, climate.mean_temp, climate.precipitation

FROM ebird_survey as ebird, covariates

WHERE climate.station_name = \`ARBORETUM UNIV WIS (WI)\` AND ebird.county = \`Dane\` AND ST_DWithin(ebird.geom, climate.geom, 1000) AND ebird.Mniotilta_varia > 0

ebird on postgres@PostgreSQL 9.6

```sql
SELECT ebird.Mniotilta_varia, climate.mean_temp, climate.precipitation
FROM ebird_survey as ebird, climate
WHERE climate.station_name ='ARBORETUM UNIV WIS (WI)' and ebird.county = 'Dane' and ST_DWithin(ebird.geom, climate.geom, 1000) and ebird.Mniotilta_varia > 0
```

Data Output | Explain | Messages | History

	mniotilta_varia numeric	mean_temp character varying (10)	precipitation character varying (10)
1	1	M	0
2	1	M	0
3	1	M	0
4	1	M	0
5	1	M	0
6	2	M	0
7	2	M	0
8	1	M	0
9	3	M	0
10	2	M	0
11	2	M	0
12	1	M	0
13	1	M	0
14	1	11.5	0
15	1	11.5	0
16	1	11.5	0
17	1	11.5	0
18	1	11.5	0
19	2	11.5	0

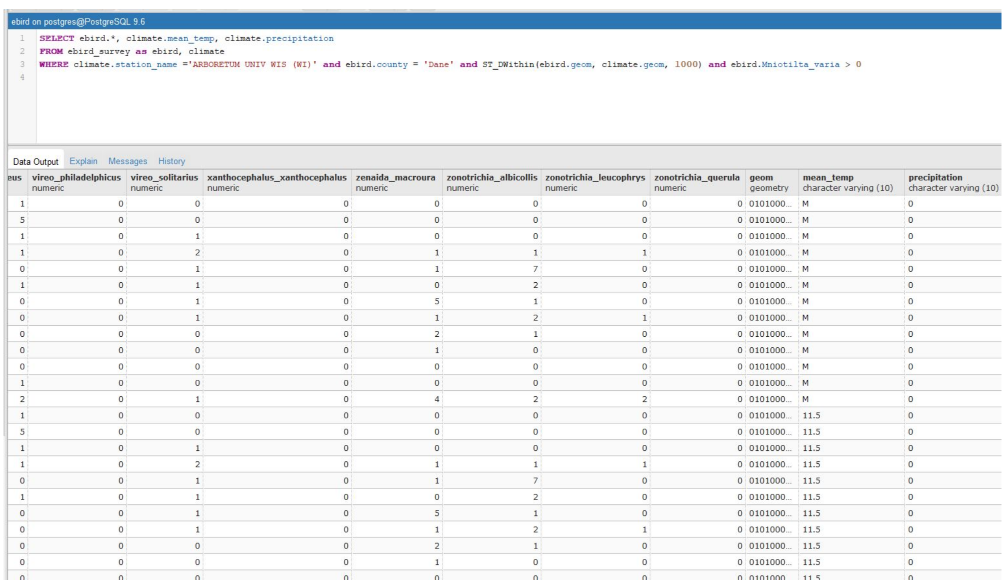
Query 4: Alongside all the information available in the eBird table, what was the precipitation amount and average daily temperature for each bird observation within 1000 meters of the University of Wisconsin Arboretum weather station?

This query was designed to look at all likely weather occurring during observations of black-and-white warblers. We chose 1000 meters as a reasonable distance that local weather within that diameter would be similar.

SELECT ebird.*, climate.mean_temp, climate.precipitation

FROM ebird_survey AS ebird, climate

WHERE climate.station_name = `ARBORETUM UNIV WIS (WI)` AND ebird.county = `Dane` AND ST_DWithin(ebird.geom, climate.geom, 1000) AND ebird.Mniotilta_varia > 0




```
SELECT ebird.*, climate.mean_temp, climate.precipitation
FROM ebird_survey as ebird, climate
WHERE climate.station_name ='ARBORETUM UNIV WIS (WI)' and ebird.county = 'Dane' and ST_DWithin(ebird.geom, climate.geom, 1000) and ebird.Mniotilta_varia > 0
```


eus	vireo_philadelphicus numeric	vireo_solitarius numeric	xanthocephalus_xanthocephalus numeric	zenaida_macroura numeric	zonotrichia_albicollis numeric	zonotrichia_leucophrys numeric	zonotrichia_querula numeric	geom geometry	mean_temp character varying (10)	precipitation character varying (10)
1	0	0	0	0	0	0	0	0101000...	M	0
5	0	0	0	0	0	0	0	0101000...	M	0
1	0	1	0	0	0	0	0	0101000...	M	0
1	0	2	0	1	1	1	0	0101000...	M	0
0	0	1	0	1	7	0	0	0101000...	M	0
1	0	1	0	0	2	0	0	0101000...	M	0
0	0	1	0	5	1	0	0	0101000...	M	0
0	0	1	0	1	2	1	0	0101000...	M	0
0	0	0	0	2	1	0	0	0101000...	M	0
0	0	0	0	1	0	0	0	0101000...	M	0
0	0	0	0	0	0	0	0	0101000...	M	0
1	0	0	0	0	0	0	0	0101000...	M	0
2	0	1	0	4	2	2	0	0101000...	M	0
1	0	0	0	0	0	0	0	0101000...	11.5	0
5	0	0	0	0	0	0	0	0101000...	11.5	0
1	0	1	0	0	0	0	0	0101000...	11.5	0
1	0	2	0	1	1	1	0	0101000...	11.5	0
0	0	1	0	1	7	0	0	0101000...	11.5	0
1	0	1	0	0	2	0	0	0101000...	11.5	0
0	0	1	0	5	1	0	0	0101000...	11.5	0
0	0	1	0	1	2	1	0	0101000...	11.5	0
0	0	0	0	2	1	0	0	0101000...	11.5	0
0	0	0	0	1	0	0	0	0101000...	11.5	0
0	0	0	0	0	0	0	0	0101000	11.5	0

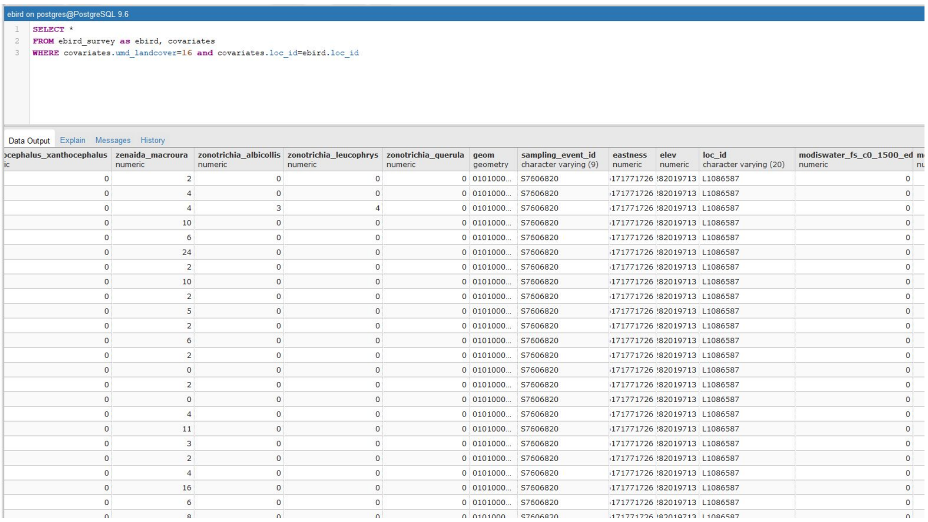
Query 5: What bird species were observed when the land cover type is classified as barren?

This query allows the user to search for birds based on a particular land cover classification code that is associated with where they were observed, such as within a deciduous forests, on cropland, grasslands, and more. This allows us to get an idea of the different species that might be associated with a particular land class type. It could also be used to sort based on other covariates like the percent of edge in an area.

SELECT *

FROM ebird_survey AS ebird, covariates

WHERE covariates.umd_landcover = 16 AND ebird.loc_id = covariates.loc_id



ebird on postgres@PostgreSQL 9.6

```
SELECT *
FROM ebird_survey as ebird, covariates
WHERE covariates.umd_landcover=16 and covariates.loc_id=ebird.loc_id
```

Data Output Explain Messages History

ocephalus_xanthocephalus ic	zenaida_macroura numeric	zonotrichia_albicollis numeric	zonotrichia_leucophrys numeric	zonotrichia_querula numeric	geom geometry	sampling_event_id character varying (9)	eastness numeric	elev numeric	loc_id character varying (20)	modiswater_fs_c0_1500_ed numeric	m nu
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	4	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	4	3	4	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	10	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	6	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	24	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	10	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	5	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	6	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	0	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	0	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	4	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	11	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	3	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	2	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	4	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	16	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	6	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	
0	8	0	0	0	0101000...	S7606820	171771726	82019713	L1086587	0	

Query 6: What land cover types are associated with a particular bird species?

When querying for a particular bird species, this allows the user to identify the habitats they are typically observed in.

SELECT ebird.aix_sponsa, ebird.year, covariates.umd_landcover, covariates.umd_watercover

FROM ebird_survey AS ebird, covariates

WHERE ebird.aix_sponsa > 0 AND covariates.loc_id = ebird.loc_id ebird on postgres@PostgreSQL 9.6

```sql
SELECT ebird.aix_sponsa, ebird.year, covariates.umd_landcover, covariates.umd_watercover
FROM ebird_survey AS ebird, covariates
WHERE ebird.aix_sponsa>1 and covariates.loc_id=ebird.loc_id
```

Data Output | Explain | Messages | History

	aix_sponsa numeric	year integer	umd_landcover integer	umd_watercover integer
1	6	2004	3	2
2	6	2004	3	2
3	6	2004	3	2
4	6	2004	3	2
5	6	2004	3	2
6	6	2004	3	2
7	6	2004	3	2
8	6	2004	3	2
9	6	2004	3	2
10	6	2004	3	2
11	6	2004	3	2
12	6	2004	3	2
13	6	2004	3	2
14	6	2004	3	2
15	6	2004	3	2
16	6	2004	3	2
17	6	2004	3	2
18	6	2004	3	2
19	6	2004	3	2
20	6	2004	3	2
21	6	2004	3	2

Conclusions

While our database only applies to Wisconsin, and more specifically to Dane County, it is quite possible for it to be expanded to entire states and countries. Databases like this are very useful for ecological research and understanding the implications of habitat and climate change on the arrival dates and detection density of birds. We produced an initial map with PostGIS that simply visualized all of the eBird surveys taken in Wisconsin during the years we included in our database (figure 3).

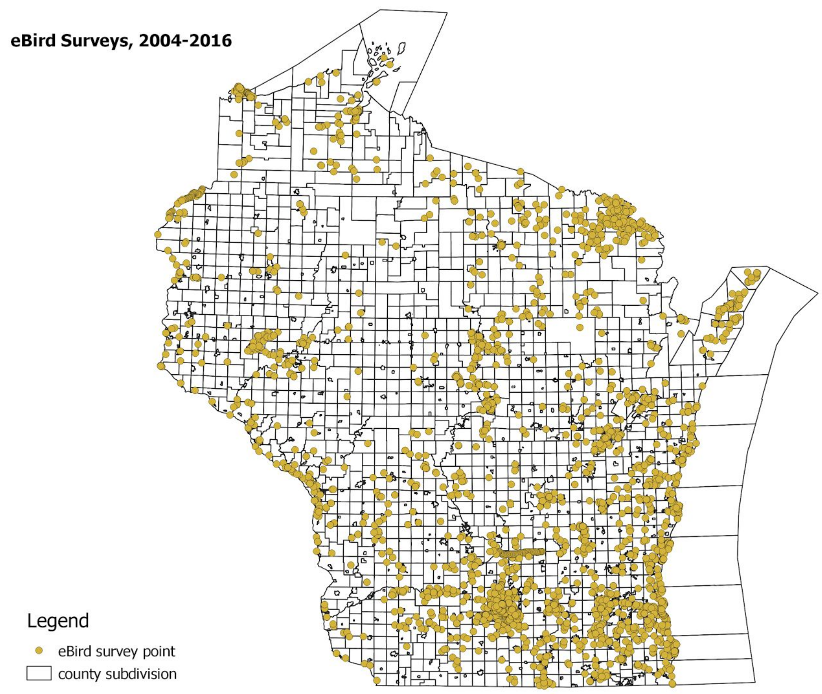


Figure 3. eBird surveys completed in Wisconsin during 2004, 2007, 2010, 2013 and 2016

With our database, it is also possible to search for all detections of any given species. For example, we created a query to search for all detections of Black -and- white Warblers (*Mniotilita varia)*. Using our query results we imported locations into QGIS, where we overlaid this new point layer with the total eBird survey points (Figure 4).

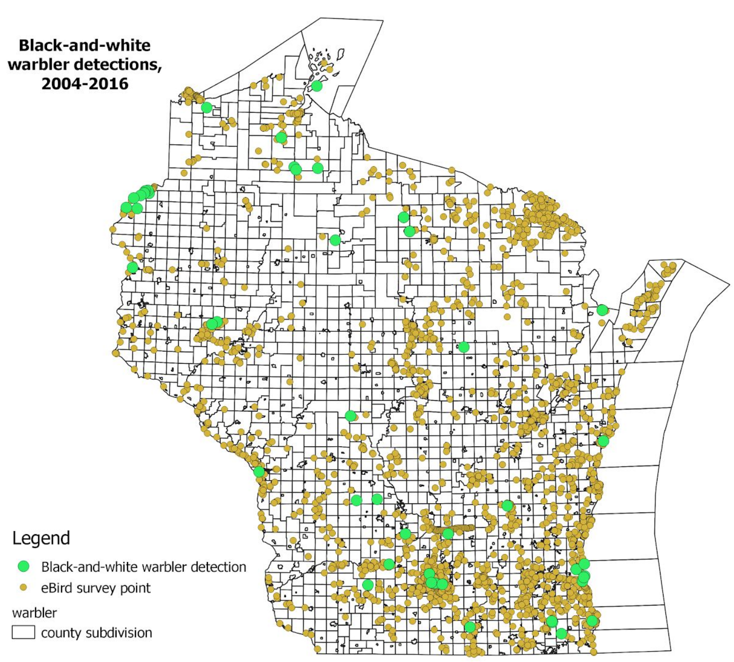


Figure 4. eBird surveys where Black-and- white Warblers were detected.

One must remember that eBird is comprised only of citizen science detection data, survey times and locations are not determined by any rigorous survey method. Some places may be visited more frequently by observers than others. However, dot density-style maps based on this data are still very useful for examining the general range of various bird species throughout Wisconsin.

Discussion

We did face a few problems in the creation and implementation of our database, most of which originated from the fact that we received a massive amount of data from eBird. With our data request we received all observations and landscape data from the western hemisphere for all years from 2002- 2016. CSV's we obtained also included all potential extant species from the western hemisphere, this is a huge amount of data. While this is broken down by year, this still led to us having some CSV files that were too large to work with in EXCEL. Even after limiting it to surveys done in Dane County we found that CSV files were too large for pgAdmin to convert to tables because files still included columns for all species in the western hemisphere (over 4,600!) and pgAdmin has a table column limit around 500. We mitigated these complications by only including surveys done in Wisconsin and cleaning remaining surveys to only include species that were likely to be found in Wisconsin, still approximately 300 species. The covariates tables were left as-is, given that there was no easy way to pare it down.

We found that even after we had limited our data we had to be very careful designing queries. We found that queries that involved joins between our climate and survey data would take an inordinate length of time that often exceeded the time we had on the computers we were using. When we wrote these queries to limit the return data they were successful. This process forced us to be more mindful about what data we actually wanted instead of just returning everything available. While we are both excited about the utility of our database as is it could provide even further application if expanded to include more data including migration lengths of species that are only seasonal in North America.

Appendix Tables:

COUNT_TYPE	Description	Type
P21	eBird – Stationary Count	STATIONARY
P22	eBird – Traveling Count	TRAVELING
P23	eBird – Exhaustive Area Count	AREAL
P34	RMBO Early Winter Waterbird Count	STATIONARY
P35	eBird – My Yard Count	STATIONARY or TRAVELING
P39	eBird Vermont – LoonWatch	AREAL
P40	My Yard eBird – Standardized Yard Count	STATIONARY
P41	eBird – Rusty Blackbird Blitz	TRAVELING
P45	eBird California – YellowBilledMagpie Traveling	TRAVELING
P46	eBird Caribbean – CWC Stationary Count	STATIONARY
P47	eBird Caribbean – CWC Area Search	AREAL
P48	eBird Random Location Count	TRAVELING
P49	eBird Peru – Coastal Shorebird Survey	STATIONARY or TRAVELING or AREAL
P50	Caribbean Martin Survey	STATIONARY
P51	Audubon NWR Protocol	TRAVELING or AREAL
P58	Texas Shorebird Survey	TRAVELING
P59	TNC California Waterbird Count	STATIONARY
P60	eBird Pelagic Protocol	TRAVELING
P61	IBA Canada	TRAVELING
P62	Historical	STATIONARY or TRAVELING or AREAL
P63	Nocturnal Count	TRAVELING
P64	Traveling – Property Specific	TRAVELING
P65	Portugal Breeding Bird Atlas	TRAVELING
P66	Birds 'n' Bogs Survey	TRAVELING
P69	California Brown Pelican Survey	TRAVELING

Appendix 1. Survey type codes. (Fink)

Class	UMD Land Cover	QA Water Value
0	Water	Shallow Ocean
1	Evergreen Needleleaf Forest	Land
2	Evergreen Broadleaf Forest	Ocean coastlines and lake shores
3	Deciduous Needleleaf Forest	Shallow inland water
4	Deciduous Broadleaf Forest	Ephemeral water
5	Mixed Forest	Deep Inland Water
6	Closed Shrublands	Moderate or continental ocean
7	Open Shrublands	Deep Ocean
8	Woody Savannas	
9	Savannas	
10	Grasslands	
12	Croplands	
13	Urban and built-up	
16	Barren	

Appendix 2. Land and Water Class codes. (Fink)

Variable Name	Comments
SAMPLING_EVENT_ID	Unique identifier for each data sample / checklist.
EASTNESS	This variable represents a combination of topographical slope and aspect. The sine of the slope is multiplied by the sine of the aspect, transforming a circular variable to a continuous [-1,1]. Derived from the 1km median of the GMTED median product (Amatulli et al. 2017). http://www.earthenv.org/topography
ELEV	One kilometer resolution digital elevation product. Point value representing the elevation in meters at the lat/long of the checklist. Derived from the 1km median GMTED median product (Amatulli et al. 2017). http://www.earthenv.org/topography
LOC_ID	Identifier for each observation location.
MODISWATER_FS_C[X]_1500_ED	Edge density for patches of water cover type [X] within a 6 × 6 neighborhood of 500 meter pixels. Ratio of total edge length to neighborhood (meters per hectare). See table 3 for definitions of the [X] types.
MODISWATER_FS_C[X]_1500_LPI	Largest patch index. Percentage of a 6 × 6 neighborhood of 500 meter pixels comprised by the largest patch of water cover of type [X].
MODISWATER_FS_C[X]_1500_PD	Patch density. Number of patches of water cover type [X] per 100 hectares within 6 × 6 neighborhood of 500 meter pixels.
MODISWATER_FS_C[X]_1500_PLAND	Percent of surrounding a 6 x 6 neighborhood of 500 meter pixels that is water cover type [X].
NORTHNESS	This variable represents a combination of topographical slope and aspect. The sine of the slope is multiplied by the cosine of the aspect, transforming a circular variable to a continuous [-1,1]. Derived from the 1km median of the GMTED median product (Amatulli et al. 2017). http://www.earthenv.org/topography
UMD_FS_C[X]_1500_ED	Edge density for patches of land cover type [X] within a 6 × 6 neighborhood of 500 meter pixels. Ratio of total edge length to neighborhood (meters per hectare).
UMD_FS_C[X]_1500_LPI	Largest patch index. Percentage of a 6 × 6 neighborhood of 500 meter pixels comprised by the largest patch of land cover type [X].
UMD_FS_C[X]_1500_PD	Patch density. Number of patches of land cover type [X] per 100 hectares within a 6 × 6 neighborhood of 500 meter pixels.
UMD_FS_C[X]_1500_PLAND	Percent of surroundings in a 6 × 6 neighborhood of 500 meter pixels that is of land cover type [X].
UMD_LANDCOVER	Class code from 500 meter resolution landcover MCD12Q1 production, using the UMD classification.
	https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1
UMD_WATERCOVER	Class code from 500 meter water cover classification derived from the MODIS Land Cover Type QA Science Data Set. https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1
YEAR	Four digit year (the same value as the four digit YEAR value found in the sampling event covariates).

Appendix 3. Extended Covariate Descriptions (Fink).

References

Atkinson CT, Utzurrum RB, LaPointe DA, Camp RJ, Crampton LH, & Foster JT, Giambelluca TW (2014). Changing climate and the altitudinal range of avian malaria in the Hawaiian Islands - an ongoing conservation crisis on the island of Kaua'i. Glob Chang Biol, 20:2426–2436.

Chase, J. M., & Leibold, M. A. (2003). *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press.

Conant, S. (1988). Saving endangered species by translocation: are we tinkering with evolution?.BioScience, 38(4), 254-257.

Crampton, L. H., Brinck, K. W., Pias, K. E., Heindl, B. A., Savre, T., Diegmann, J. S., & Paxton, E. H. (2017). Linking occupancy surveys with habitat characteristics to estimate abundance and distribution in an endangered cryptic bird. Biodiversity and Conservation, 26(7), 1525-1539.

Fink, D., Auer, T., Obregon, F., Hochachka, W.M., Iliff, M., Sullivan, B., Wood, C., Davies, I., & Kelling, S. (2018). The ebird reference dataset, version 2016. *Cornell Lab of Ornithology and National Audubon Society, Ithaca, NY*.