

Project Tidy Data

David Vasquez, Tomoki Takeuchi, John Hong, Jose Hernandez

10/30/2021

Data Introduction

Our project partner is Professor Proppe, who is doing research on golden-cheeked warblers in Wild Basin Wilderness Preserve, a constituent of the larger Balcones Canyonlands Preserve (BCP) system. The golden-cheeked-warbler is an endangered bird species in Texas, with very narrow habitat in central Texas. The ultimate goal is to preserve this species in central Texas and prevent them from going extinct in Texas. He is seeking to investigate the effects that urban development, highways and housing developments in particular, may be having on the behavior of the Golden-cheeked Warblers as Austin develops into west, closer to the habitat of the warblers. The mission is to determine whether if there is any trend in the warblers within the reserve in response to development around the reserve.

Questions:

- 1) Are there things that are driving the golden-cheeked warbler out of the area (housing, roads, etc.)?
- 2) Are golden-cheeked warblers moving west away from development?
- 3) How do spatial trends change over time?

To answer these questions, we will first need to find the data about urban development. We can combine the data with the shapefiles of birds observations. One idea that we have is making a grid. In this way, we can compute the number of dots (birds observation) in each grid, and how it changed over time. In addition to making grids, we could also use longitude and latitude as response variables to study where their habitat has been shifting over years.

Data Processing

- 1) Retrieved shape file data for golden-cheeked warbler (GCW) observations and BCP system boundaries from Dr. Darren Proppe
- 2) Using ArcGIS Pro, filtered the boundaries to include only the following properties:
 - Barton Creek
 - Double J&T
 - Emma Bike Park
 - Emma Extension
 - Forest Ridge
 - Hamilton
 - Kent Butler
 - Reicher Ranch
 - Vireo/Wild Basin

- Vista Point
- 3) **Filtered the GCW observations to include only observations collected by specific survey types per each year:**
 - 2011 - 100 Acre Plot
 - 2012 - Spot Mapping
 - 2013 - Intensive Study Plot and Vickery
 - 2014 - Intensive Study Plot
 - 2015 - Intensive Study Plot
 - 2016 - Intensive Study Plot
 - 2017 - Intensive Study Plot
 - 2018 - Intensive Study Plot and 100 Acre Plot
 - 2019 - Intensive Study Plot
 - 4) **Created a grid using ‘Create Fishnet’ for the filtered boundaries with 100-meter x 100-meter cells**
 - 5) **Counted the GCW observations per cell per year using ‘Summarize Within’**
 - 6) **Joined the GCW observations for each year into one grid using ‘Add Join’**
 - 7) **Built 1-kilometer buffers around each cell using ‘Buffer’**
 - 8) **Downloaded the National Land Cover Database (NLCD) impervious surface raster for all years (30-meter resolution) for the study area**
 - 9) **Clipped the NLCD raster by the buffer for each cell using ‘Clip Raster’**
 - 10) **Converted the NLCD raster to a point vector using ‘Raster to Point’**
 - 11) **Calculated the mean impervious surface within each buffer (mean of 30 m cells) for each relevant year (2011, 2013, 2016, 2019) using ‘Zonal Statistics’**
 - 12) **Joined the mean impervious surface value for each buffer to the associated cell in the grid using ‘Add Join’**
 - 13) **Filtered the grid to include GCW observations for each cell for only relevant years (2011, 2013, 2016, 2019)**
 - 14) **Added UTM Easting and UTM Northing coordinates for each cell centroid using ‘Add Geometry Attributes’**
 - 15) **Added the appropriate property name manually to each cell based on the filtered plot boundaries**
 - 16) **Converted the grid attribute table to an Excel file using ‘Table to Excel’**
 - 17) **Modified the Excel file for use in RStudio, creating a row for each cell and specific year**

Reading in Dataset

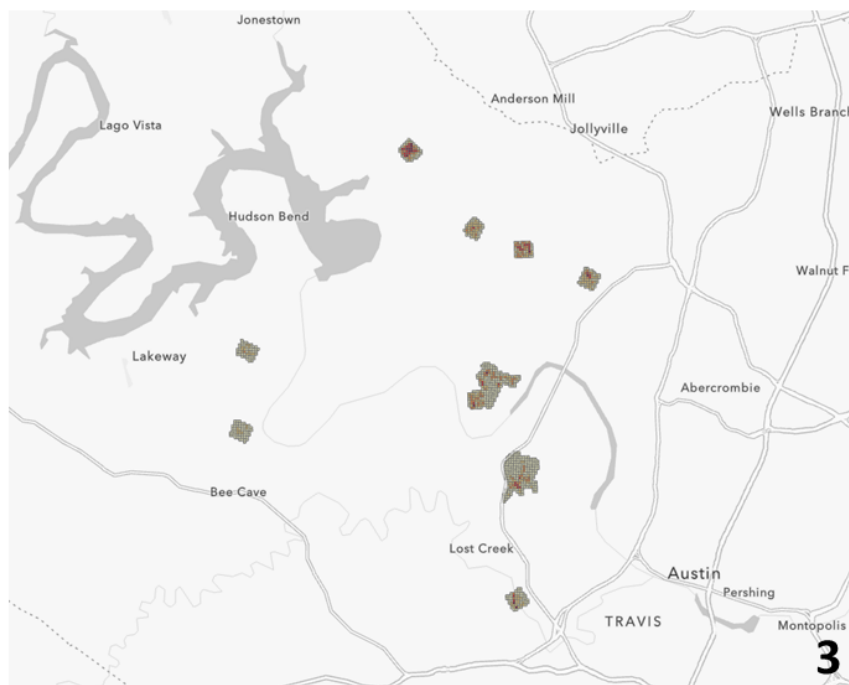
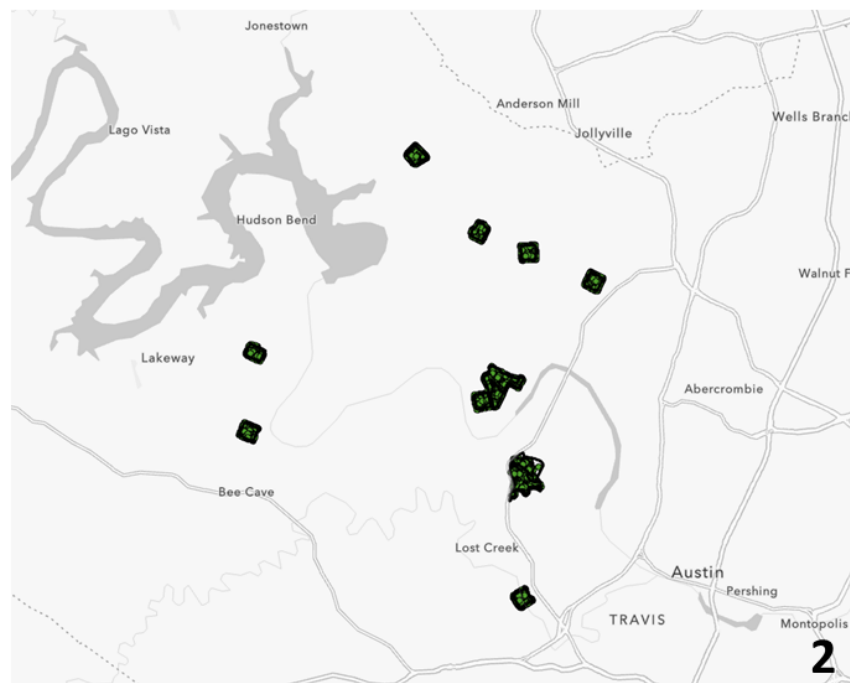
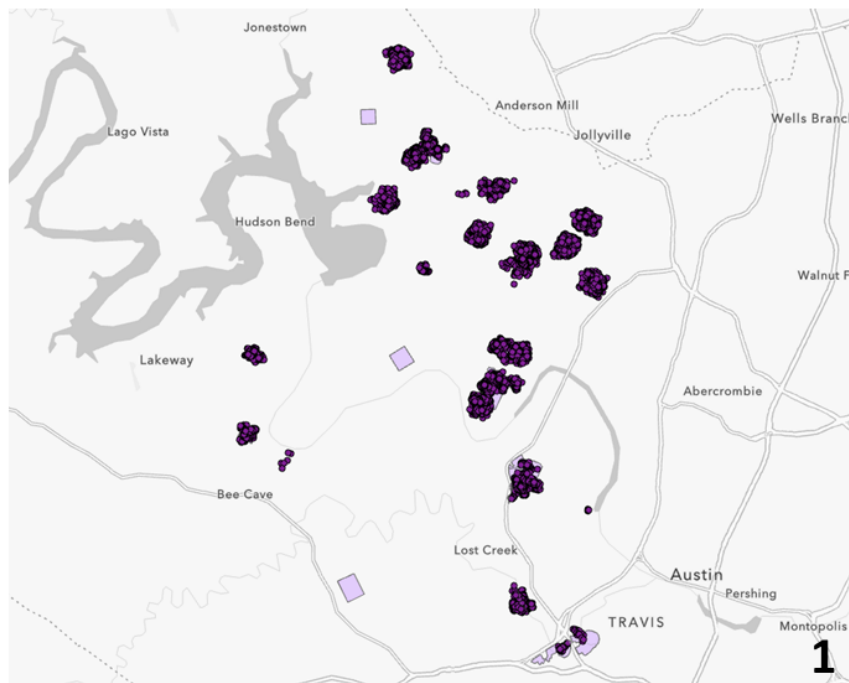
```
warblers <- read_csv('https://docs.google.com/spreadsheets/d/e/2PACX-1vQwR56LrrsuT8BMSLbN6bmu8CPLIFhX5L...')

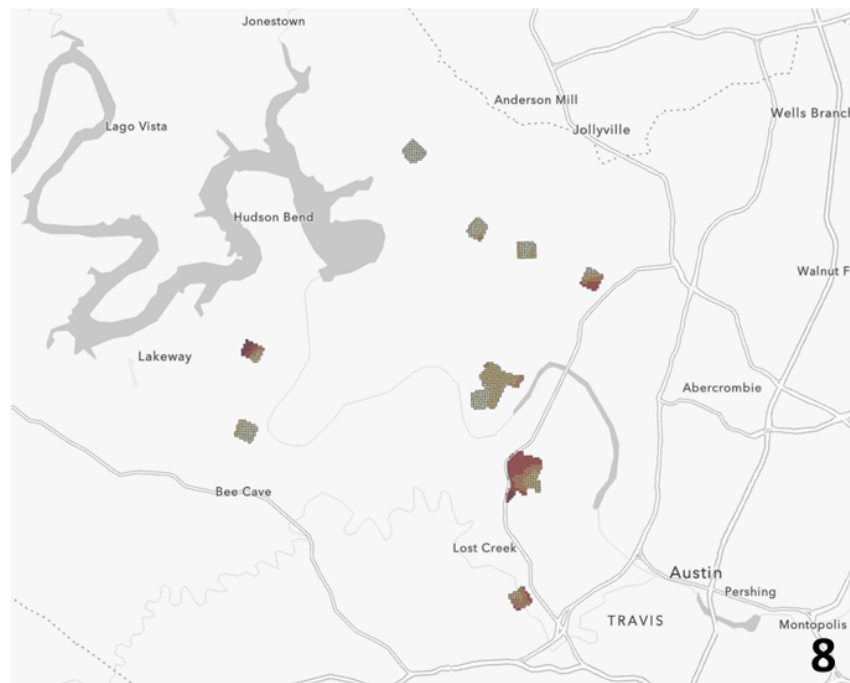
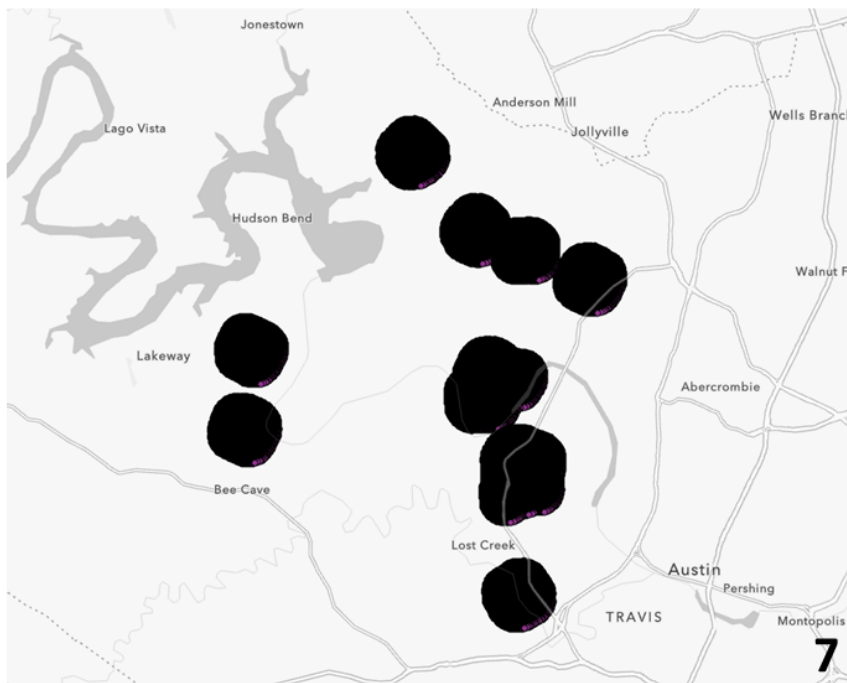
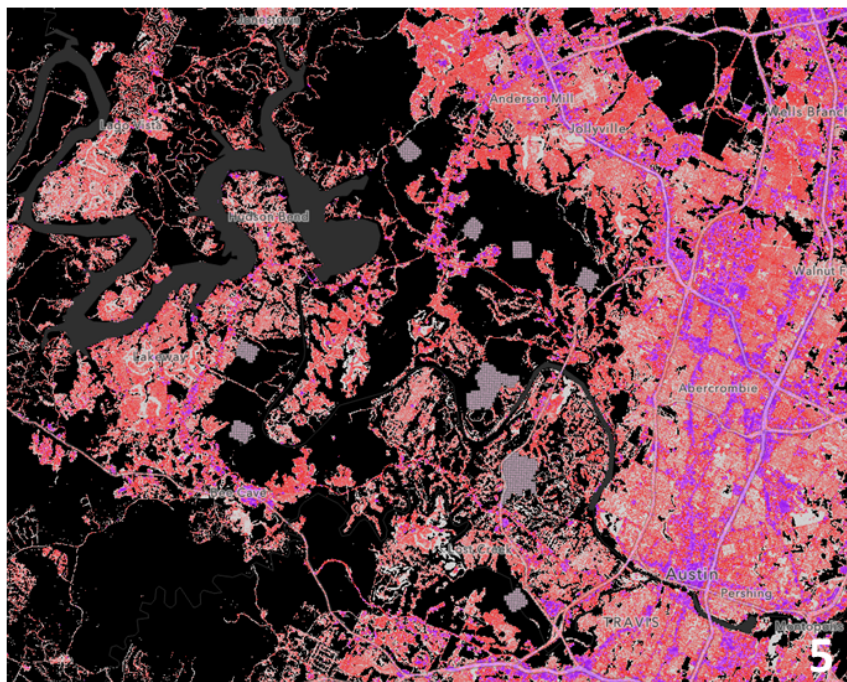
## Rows: 3452 Columns: 9
## -- Column specification -----
## Delimiter: ","
```

```

## chr (1): Property
## dbl (7): GCW, Year, IMPCELLS, MeanImp, CENTROID_X, UTM_E, UTM_N
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
warblers <- warblers %>%
  mutate(
    UTM_E = UTM_E / 1000000,
    UTM_N = UTM_N / 1000000,
    Year = as.factor(Year),
    GCW = as.factor(GCW)
  )
# Divided UTM_E and UTM_N by 1,000,000 to fix scaling error

```





Final_Grid

Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy

FID	Shape *	Points2011	Points2013	Points2016	Points2019	Point_Count	MEAN_grid_	Point_Co_1	MEAN_grid1	Point_Co_2	MEAN
1	Polygon	0	1	0	0	3949	2.482147	3949	2.482147	3949	2
9	Polygon	0	6	0	0	3948	3.112209	3948	3.125633	3948	1
16	Polygon	0	5	0	0	3947	3.195085	3947	3.208513	3947	3
43	Polygon	0	1	0	5	3950	0.878228	3950	0.889114	3950	0
50	Polygon	0	12	0	1	3944	1.155933	3944	1.190416	3944	1
51	Polygon	0	2	0	1	3947	1.100836	3947	1.135293	3947	1
56	Polygon	0	7	0	0	3952	1.413209	3952	1.447621	3952	1
73	Polygon	0	1	1	0	3950	4.600253	3950	4.637215	3950	5
104	Polygon	0	2	0	1	3947	3.353686	3947	3.358753	3947	5
110	Polygon	0	1	0	1	3952	5.638664	3952	5.648785	3952	7
137	Polygon	0	2	0	0	3946	4.267866	3946	4.267866	3946	4
170	Polygon	0	3	1	0	3957	2.840788	3957	2.864038	3957	2
171	Polygon	0	3	0	0	3953	2.977232	3953	3.000506	3953	3
255	Polygon	0	1	0	0	3948	7.257092	3948	7.445035	3948	7
276	Polygon	0	1	0	0	3951	10.328778	3951	10.551506	3951	1
307	Polygon	0	1	0	4	3943	4.670555	3943	4.791276	3943	4
310	Polygon	0	3	0	1	3947	5.345326	3947	5.387129	3947	5
319	Polygon	0	1	0	2	3952	5.089069	3952	5.24747	3952	5
320	Polygon	0	2	0	2	3951	4.769172	3951	4.794482	3951	4
321	Polygon	0	43	6	31	3947	4.868001	3947	4.909805	3947	5
332	Polygon	0	3	5	1	3951	5.270058	3951	5.440395	3951	1
334	Polygon	0	3	5	6	3948	4.489615	3948	4.531408	3948	4
351	Polygon	0	3	0	0	3944	6.284736	3944	6.50431	3944	6
357	Polygon	0	2	1	1	3950	4.352152	3950	4.519494	3950	4
360	Polygon	0	7	2	0	3950	5.383797	3950	5.541013	3950	5
377	Polygon	0	6	0	0	3950	4.936962	3950	5.131646	3950	5
378	Polygon	0	15	4	2	3945	5.445627	3945	5.640558	3945	9

9

FinalDataSetUpdated

XLSX

File Edit View Insert Format Data Tools Help

Last edit was on November 19

100% \$ % .00 123 Arial 10 B I U A

A1

FID

	A	B	C	D	E	F	G	H	I	J	K
1	FID	GCW2011	GCW2013	GCW2016	GCW2019	IMPPPOINT2011	MeanImp2011	IMPPPOINT2013	MeanImp2013	IMPPPOINT2016	MeanImp2016
6	4	0	0	0	0	3945	3.053231939	3945	3.066969687	3945	3.08643853
7	5	13	11	1	2	3946	2.697161683	3946	2.697161683	3946	2.716928535
8	6	31	39	0	15	3950	2.438987342	3950	2.438987342	3950	2.448101296
9	7	51	28	1	16	3942	2.244799594	3942	2.244799594	3942	2.253932014
10	8	3	9	12	8	3951	2.12629714	3951	2.12629714	3951	2.135408757
11	9	0	6	0	0	3948	3.112208713	3948	3.125633232	3948	3.145390071
12	10	18	50	0	4	3950	3.047594937	3950	3.061012658	3950	3.080759494
13	11	39	46	8	24	3952	2.902327935	3952	2.915738866	3952	2.935475709
14	12	34	23	5	30	3951	2.367501898	3951	2.367501898	3951	2.376613516
15	13	40	26	2	22	3944	2.207910751	3944	2.207910751	3944	2.21703854
16	14	16	7	9	14	3949	1.944542922	3949	1.944542922	3949	1.953659154
17	15	13	3	14	2	3953	1.480647609	3953	1.480647609	3953	1.489754617
18	16	0	5	0	0	3947	3.195084875	3947	3.208512795	3947	3.228274639
19	17	27	32	0	8	3953	3.158360739	3953	3.171768277	3953	3.191500126
20	18	37	43	3	15	3953	3.056665823	3953	3.070073362	3953	3.089805211
21	19	53	29	4	28	3941	2.917279878	3941	2.930728242	3941	2.950520173
22	20	43	25	3	20	3948	2.463272543	3948	2.476697062	3948	2.496453901
23	21	41	46	1	28	3947	2.04281733	3947	2.04281733	3947	2.051938181
24	22	23	27	2	17	3953	1.689096888	3953	1.689096888	3953	1.698203896
25	23	19	14	4	14	3948	1.275329281	3948	1.275329281	3948	1.284447822
26	24	1	0	1	2	3949	1.023297037	3949	1.023297037	3949	1.032413269
27	25	1	22	0	0	3937	3.180848362	3937	3.194310389	3937	3.214122428
28	26	24	44	0	27	3955	3.05335019	3955	3.066750948	3955	3.086472819
29	27	26	6	2	33	3954	2.981790592	3954	2.995194744	3954	3.014921598
30	28	28	20	2	28	3949	2.873892125	3949	2.887313244	3949	2.90708508
31	29	58	35	3	28	3950	2.387848101	3950	2.401265823	3950	2.421012658
32	30	40	22	4	15	3944	1.74137931	3944	1.74137931	3944	1.750507099
33	31	39	60	3	18	3949	1.436819448	3949	1.436819448	3949	1.445503568
34	32	32	40	2	10	3952	1.177631579	3952	1.177631579	3952	1.186740891
35	33	19	3	3	6	3953	1.047811789	3953	1.047811789	3953	1.056918796
36	34	0	0	0	0	3947	0.7572840132	3947	0.7572840132	3947	0.759564226
37	35	18	0	2	10	3950	2.999240506	3950	3.012658228	3950	3.032405083
38	36	39	2	2	21	3951	2.8423184	3951	2.855732726	3951	2.875474563
39	37	31	5	3	18	3946	2.581348201	3946	2.594779524	3946	2.614546376

+

OriginalGISOutput

FinalDataSet

Metadata

10

10