parsing_ebird_for_maxent

March 24, 2017

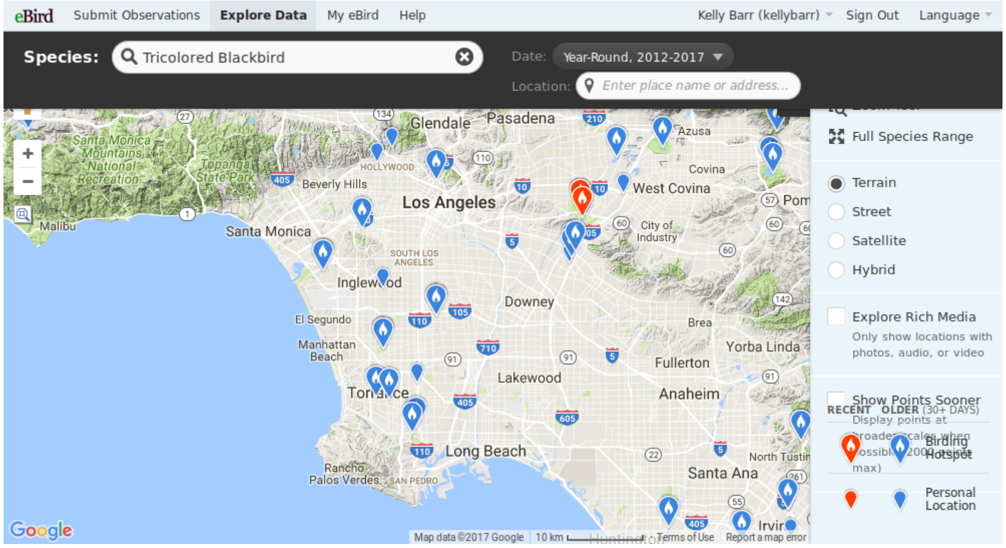
1 Mining eBird data for habitat modeling.

1.1 Kelly Barr

The Tricolored Blackbird is a declining species that is protected under the California Endangered Species Act (Meese et al. 2014).

title

An example of the data available via ebird (Sullivan et al. 2009):



Species: Tricolored Blackbird

Date: Year-Round, 2012-2017

Location: Enter place name or address...

○ Terrain
○ Street
○ Satellite
○ Hybrid

□ Explore Rich Media
Only show locations with photos, audio, or video

□ Show Points Sooner
Display points at broader scales when possible (200 points max)

TITLE

What this code does:

2 1. Parse out data of interest: dates & locations

3 2. Subset these by year

4 3. Create input files with randomized locations

```
In [2]: #This code takes in a file downloaded from ebird.  Data are behind a sign-i
        #available through wget or curl.  This code reads in the data and creates c
        #for analyses downstream.  Namely, the locations and dates for observations

        #this parser will return a list that can be used later as a combined dictio
        def parse_ebird_todict(file_name):
            import csv
            f = open(file_name,'r')
            data = csv.reader(f,delimiter='\t')
            year_lat_long = []
            for row in data:
                year_lat_long.append(row[24].split('-')[0] + ', ' + row[22] + ', '
            #there is some crap at the beginning and end of these lists to remove
            year_lat_long = year_lat_long[1:-1]
            bird_data_dict = dict()
            for line in year_lat_long:
                if line.split(', ')[0] in bird_data_dict:
                # append the new number to the existing array at this slot
                    bird_data_dict[line.split(', ')[0]].append(line.split(', ')[1]
                else:
                    # create a new array in this slot
                    bird_data_dict[line.split(', ')[0]] = [line.split(', ')[1] + ('
            f.close()
            return bird_data_dict
```

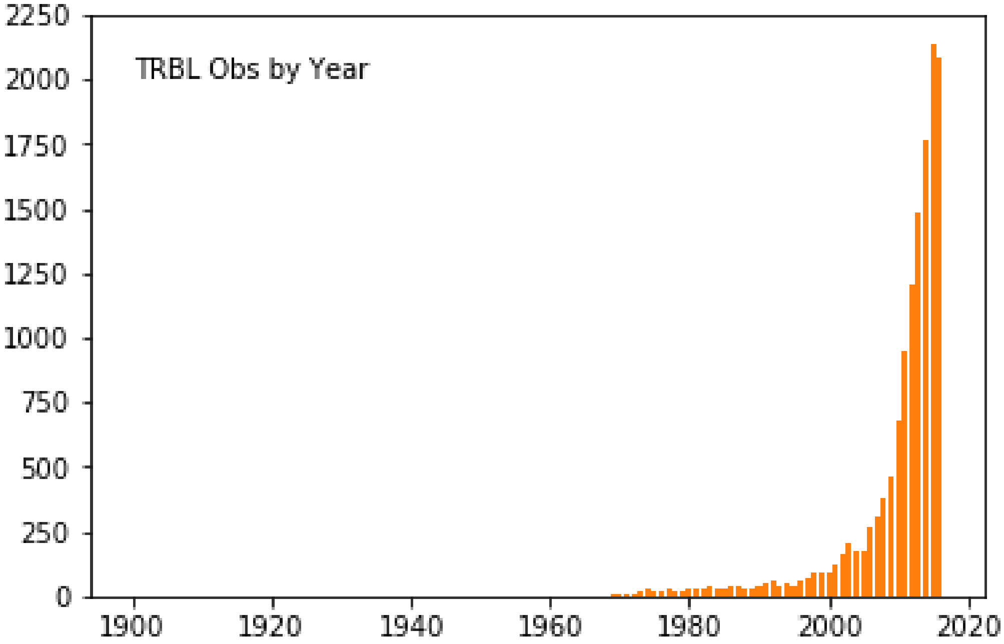
```
In [3]: trbl_example = open("trbl_ebird.txt",'r').readlines()
        #remove quotes to show what the data look like
        #trbl_example
```

```
In [4]: #making dictionary for Tricolored Blackbirds
        trbl_data = parse_ebird_todict("trbl_ebird.txt")
```

```
In [5]: #making list of years vs number of datapoints for graphics
        years = []
        number_obs = []
        for keys,values in trbl_data.items():
            years.append(int(keys))
            number_obs.append(len(trbl_data[keys])/2)
```

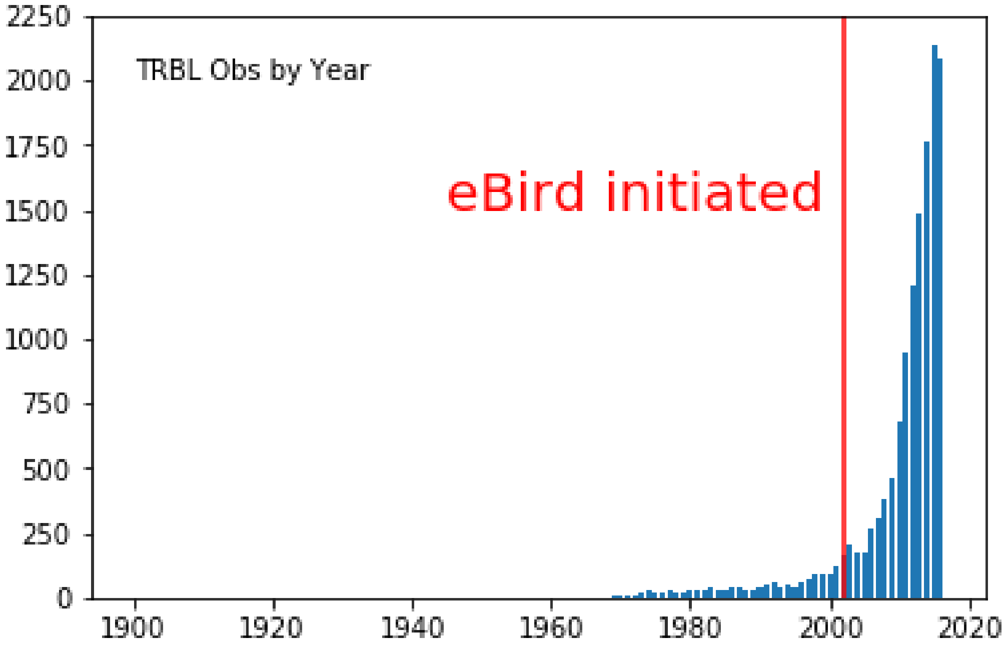
```
In [7]: import numpy as np
        import matplotlib.pyplot as plt
```

```
plt.bar(years, number_obs)
plt.text(1900, 2000, "TRBL Obs by Year",fontsize=10)
plt.show()
```



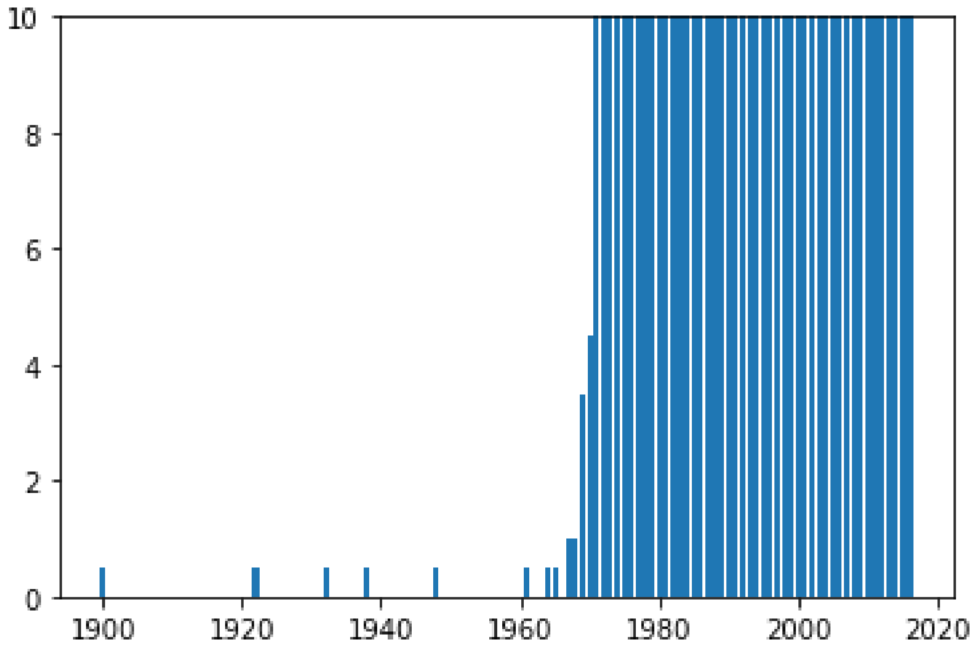
4.0.1 That's odd. eBird is a website and there are data since 1900. Let's explore that:

```
In [12]: plt.bar(years, number_obs)
         plt.axvline(x = 2002, color = "red")
         plt.text(1900, 2000, "TRBL Obs by Year",fontsize=10)
         plt.text(1945, 1500, "eBird initiated",fontsize=20, color = "red")
         plt.show()

         #this plot shows my data
```



Why do those earlier years appear anyway? Let's explore that:

```
In [29]: plt.bar(years, number_obs)
         plt.ylim(0,10)
         plt.show()

         #plot illustrates why some data need to come out
```



4.0.2 1. Parse out data of interest: dates & locations

5 2. Subset these by year

5.0.1 3. Create input files with randomized locations

```
In [11]: #this function grabs the geo data out by year
         def geos_by_year(year1,year2,bird_dict):
             list_of_years = []
             list_of_years.append(year1)
             i = 0
         #populate a list of years
             while i < (year2 - year1):
                 list_of_years.append(year1 + i)
                 i = i + 1
         #this code removes the quotes from the years so that the list of years ca
             bird_dict2 = {eval(k):v for k,v in bird_dict.items()}
             list_of_locations = []
             for key, value in bird_dict2.items():
         #conditional:  if key is in list of years
                 if key in list_of_years:
                     list_of_locations.append(value)
         #This next code combines the list 0f lists into one (flattens it).
             geos = [item for sublist in list_of_locations for item in sublist]
             return geos
```

```
In [12]: #calling function
         trbls = geos_by_year(2003,2014,trbl_data)
         j = 0
         #population second variable with data so they can be shuffled
         trbls2 = []
         while j < len(trbls):
             trbls2.append(trbls[j:j+1])
             j+=1
```

5.0.2 1. Parse out data of interest: dates & locations

5.0.3 2. Subset these by year

6 3. Create input files with randomized locations

```
In [18]: #generating files with random datapoints; 70% for model-training, 30% for
         import random
         #randomizing data
         random.shuffle(trbls2)

         outfile_training = open("training_dataset.txt",'w')
         outfile_testing = open("testing_dataset.txt",'w')
         #populating variables with subsets
         training_set = trbls[0:round(len(trbls)*0.7)]
         test_set = trbls[round(len(trbls)*0.7):]
         i = 0
         j = 0
         for line in training_set:
             outfile_training.write("TRBL" + ', ' + line + '\n')
             i += 1
         for line in test_set:
             outfile_testing.write("TRBL" + ', ' + line + '\n')
             j += 1

         outfile_training.close()
         outfile_testing.close()

In [25]: trbl_outfile = open("training_dataset.txt",'r').readlines()
         #remove hash to show data
         #trbl_outfile
```

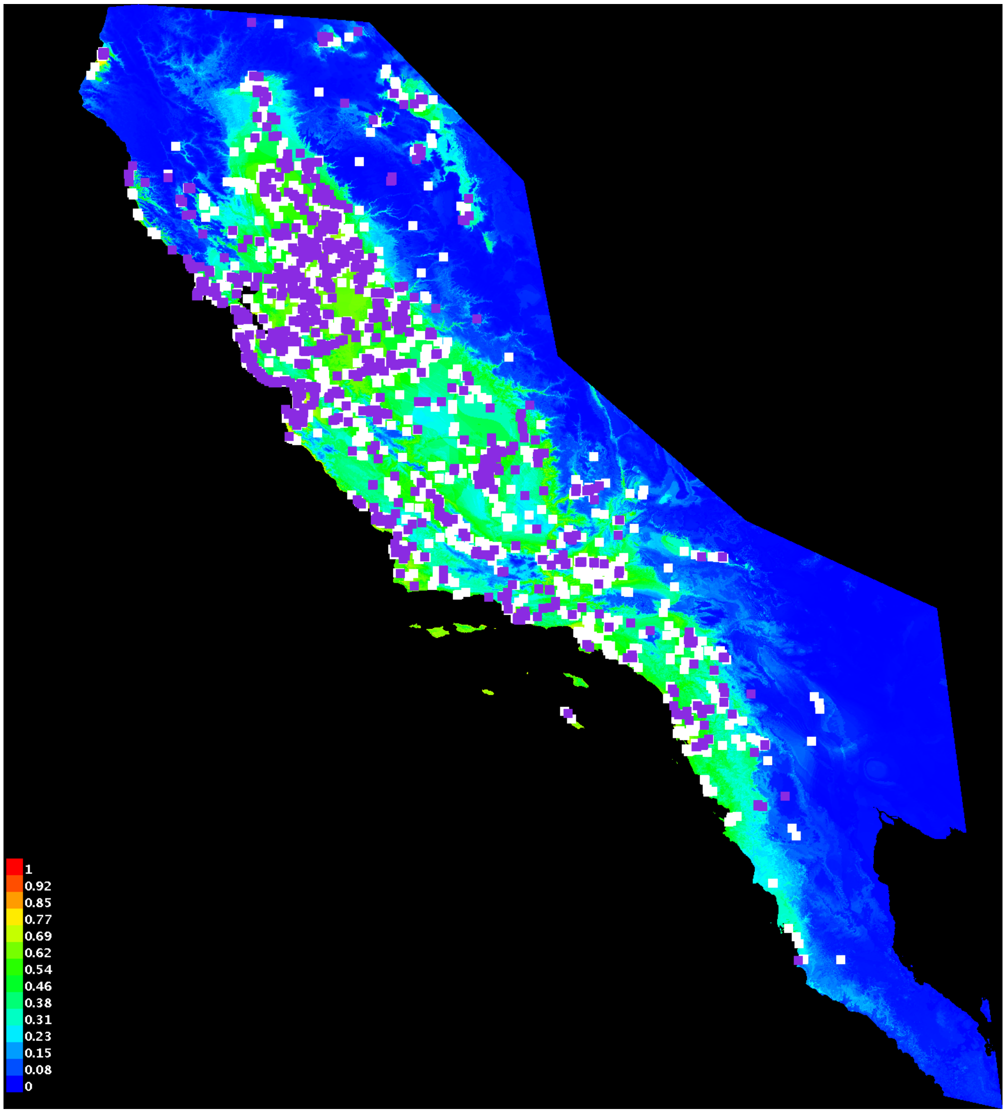
Using the program MAXENT (Phillips & Dudik 2008), I built a habitat model for the TRBL. The following is the known range of the TRBL. So this habitat model is pretty good.

6.1 References

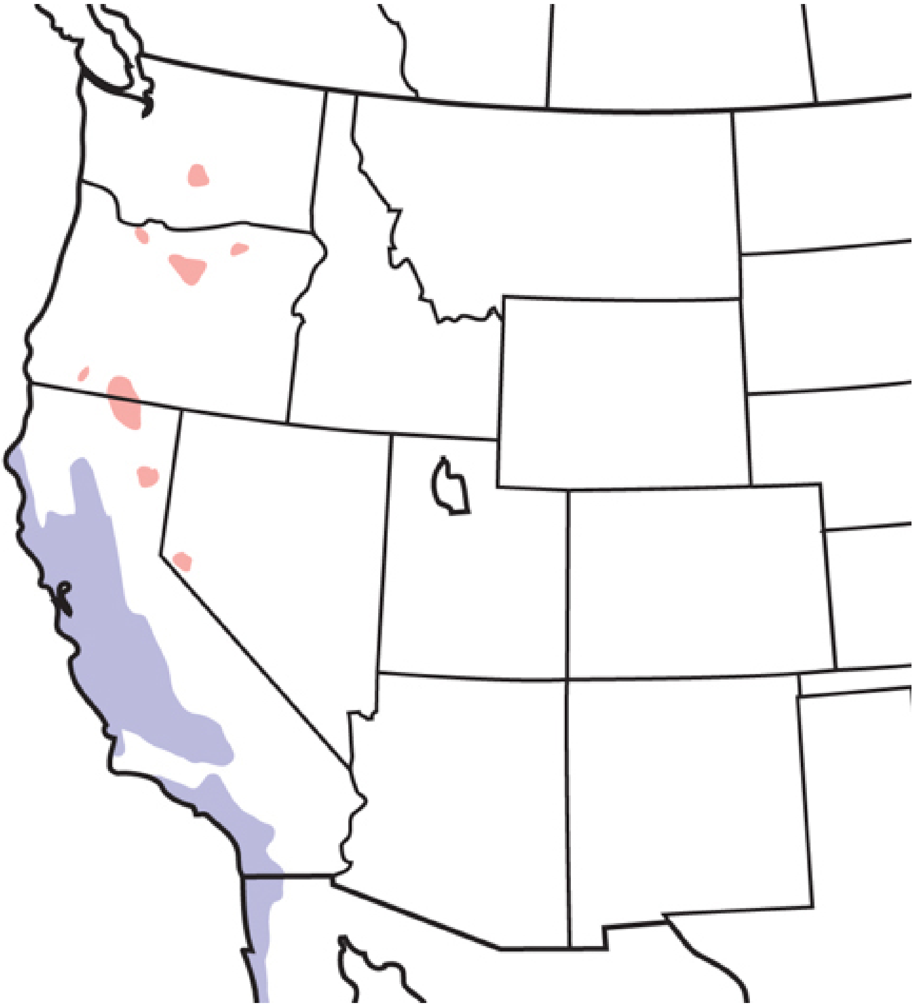
Meese, Robert J., Edward C. Beedy and William J. Hamilton, III. (2014). Tricolored Blackbird (Agelaius tricolor), The Birds of North America (P. G. Rodewald, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America: https://birdsna.org/Species-Account/bna/species/tribla

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31(2), 161-175.

Sullivan, B.L., C.L. Wood, M.J. Iliff, R.E. Bonney, D. Fink, and S. Kelling (2009) eBird: a citizen-based bird observation network in the biological sciences. Biological Conservation 142: 2282-2292.



TITLE



TITLE