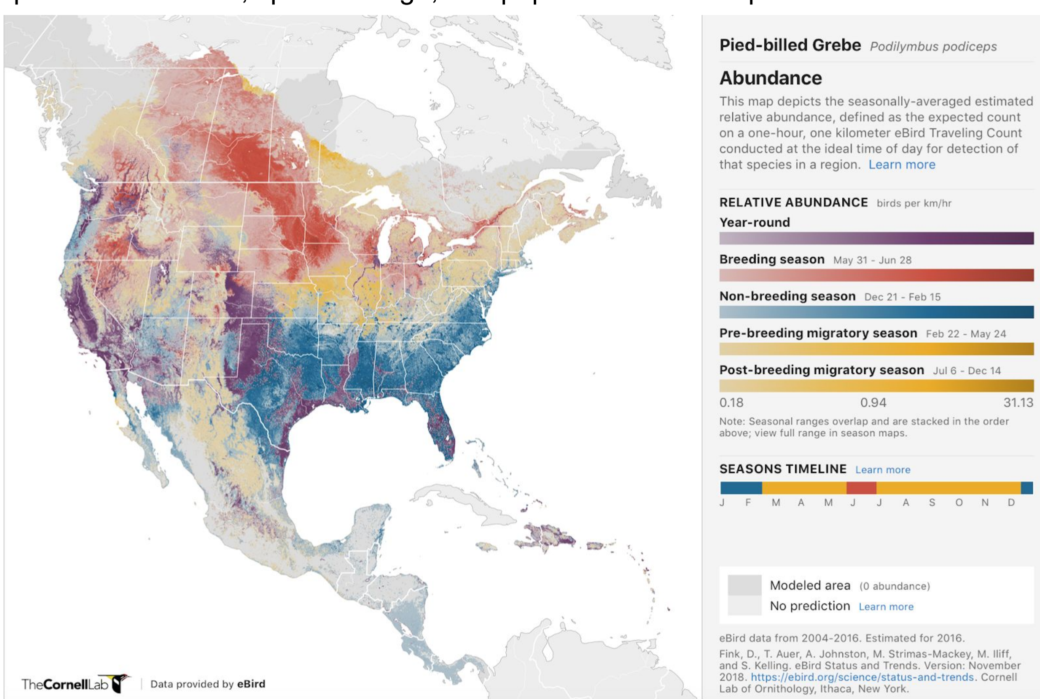
Bird Population and Habitat Visualization
Corin Thummel, Will Richards

Project Process Book

- Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.
    - We wanted to choose a project that would shed light on a topical environmental issue. The world is seeing a decline in biodiversity, with many species becoming endangered or extinct. We chose to focus specifically on birds, as they are an important part of the global ecosystem. Bird populations have been well studied, and there are many ongoing efforts to protect and monitor the declining population of birds. Avian prosperity is a good indicator for overall ecosystem health, as many birds are keystone species across a wide range of biomes. Both team members have a personal interest in birds and birding. While the Ebird.org website provides great visualizations for individual species, it does not provide any interaction or way to compare across species. We seek to make these visualizations interactive, and allow for species comparisons and data exploration.

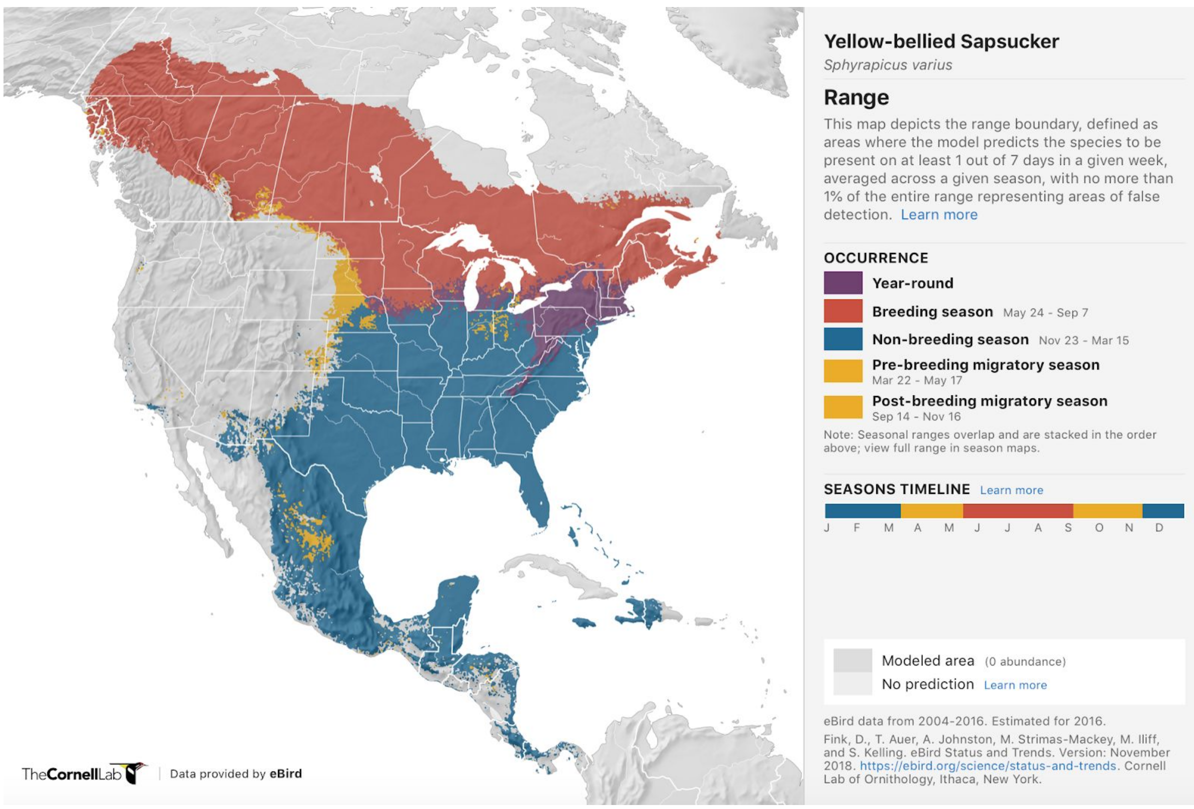
- Related Work: Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.
    - Our inspiration for this project came from the Cornell Lab of Ornithology. This lab has generated non-interactive plots to show various trends for different species of birds. Specifically, we were inspired to implement our own interactive versions of species abundance, species range, and population trend maps.



Bird Population and Habitat Visualization

Corin Thummel, Will Richards

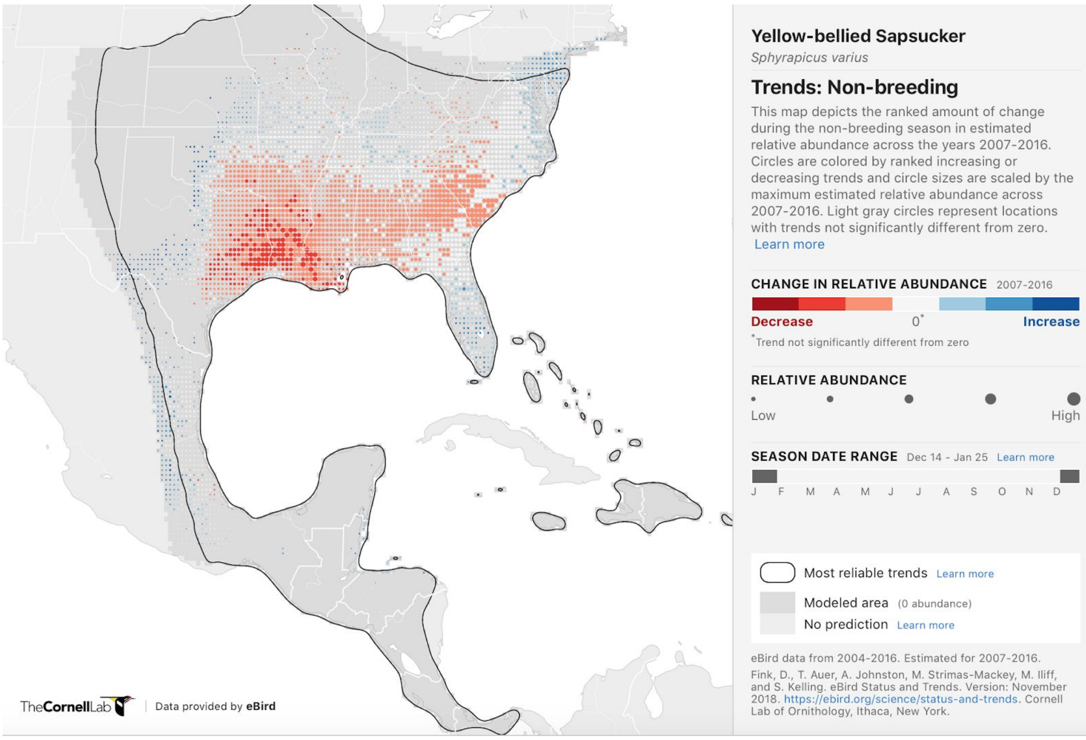
- This abundance map served as the inspiration for our first and most important abundance visualization. This map encodes the estimated relative abundance of the species at various 2.5km^2 grid cells as a heat map.
- We would like to show abundance in two ways. The first way will be generating the heatmap seen above from an R package provided with the Ebird dataset. Because this abundance data was generated from data spanning many years, the abundance heatmap only predicts abundance for 2018. We would like to have a temporal interaction where the user can inspect how abundance changes over time. To do this, we implemented a dot-density plot overlaid on a map of North america. We have raw observation counts for each species, and can plot these observations by year using a year-slider as input control.



We also plan on implementing a version of this range map. This shows the area encompassing all known observations of a species. We have also implemented a seasons timeline selector. This season slider will pass the range of dates to use for the range raster once we have our R pipeline integrated into our visualization.

Bird Population and Habitat Visualization

Corin Thummel, Will Richards



The visualization above was our inspiration for showing population trend data of a species. In the example above, trend points describe the significance in abundance change from 2007-2016. We can generate this trend data from the provided R packages. We are not sure if we want to encode trend data as P-values. We may instead decide to show the percent change in observations from 2007-2016. This is more intuitive, but does not take into account how many observation reports were collected for each year.

- Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?
  - The primary question we seek to answer is how is the overall bird population and habitat changing over time. A secondary question is to see how these trends vary amongst, and see which species are experiencing the most significant population and habitat decrease. Another secondary question we seek to answer is how bird species migration patterns are changing over time. We initially sought to answer how trends compared across bird types, but felt this would interfere with the clarity

and impact of our visualizations. This will further be discussed in the 'exploratory data analysis' section.

- Data: Source, scraping method, cleanup, etc.
  - Observational data (CSV)

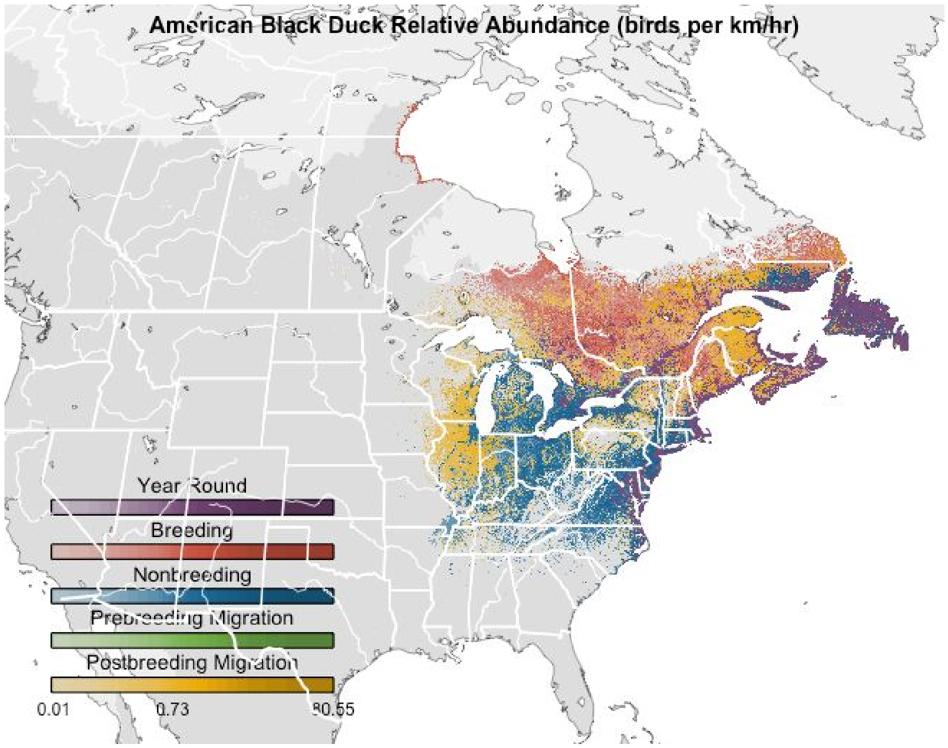
    The observational data is stored in a zipped tsv. For our initial prototype, we pulled 500 lines from the tsv and stored it as a .txt file in our project data folder. We can use d3.dsv to directly load the tsv into an array of json, pulling only the attributes that we desire from the tsv. Unfortunately, we will not be able to use this method to load observational data for our final project. If we decide to include observational data, we will need to either generate small sample files for a subset of species, or host our data in sql database. If we can generate small files enough for a responsive app, we will do that. Otherwise, we will host the observational data in an sql database. We can use node.js and MySQL to transfer data from the server to the browser. The format of the data itself is very suitable for our needs.
  - Range/Abundance and Trend data (hosted R data):

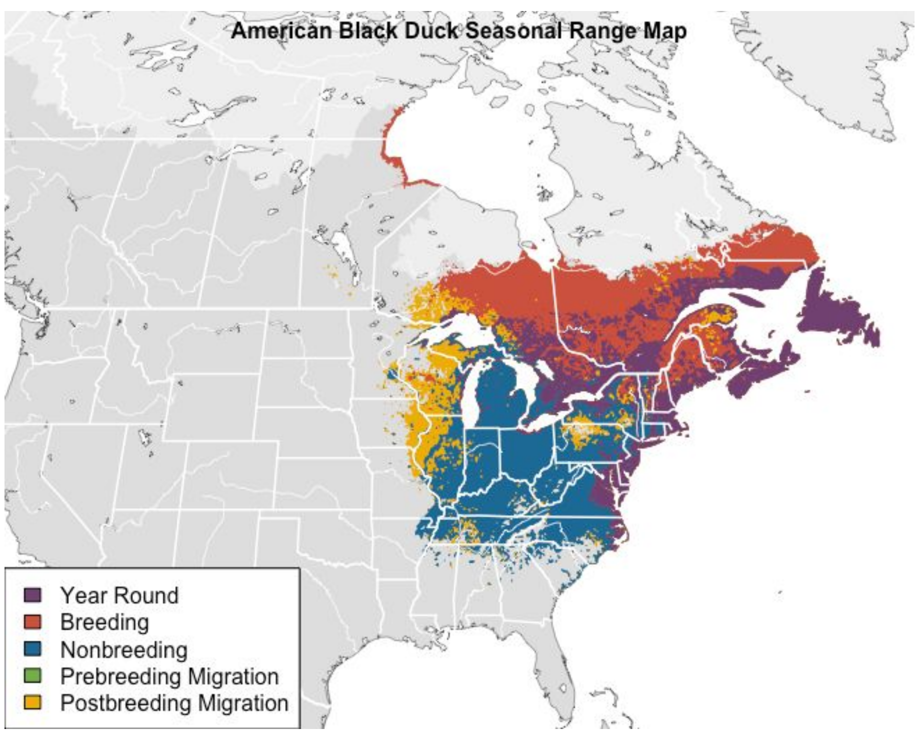
    The R data comes from an ebirds.org raster that that they provide through AWS. This data is based off the abundance data from before but they have run it through their own statistical pipeline in order to up sample bird populations in areas where fewer people travel. They reference a variety of papers but the basic idea is creating many overlapping rectangles of bird sightings and averaging out the overlaps to create an area that better captures the likelihood of seeing a bird in a given location. The data is provided to us in a .tif file that contains 52 layer raster. Each layer corresponds to a week in the year. Once we precompute some of these maps in R, we should be able to show certain layers to highlight seasonal changes.

- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
  - As mentioned above, we were initially planning on adding interactivity to the maps shown on ebirds.org by adding seasonal and yearly sliders to allow for a more in depth exploration of bird populations. Unfortunately, the rasters provided to generate these maps are not set up to work across years. They used their historical data to generate the likelihood of spotting birds in a given location but the raster itself is only for 2016. This means that when displaying these in depth maps, we will turn off the year slider and only allow seasonal brushing. Another issue is that it is difficult to display many birds at the same time because the data has such high resolution. Since each bird won't necessarily share territory, the heatmaps can get

overwhelming quickly and basically everywhere is highlighted. Therefore we suspect that we can only show one or two birds at a time and not whole categories.

- The dot density plot showed us how frequency of reports has consistently increased. We now know we need to normalize the dot density plot by report activity.
- The following two maps are examples of the R data.





This is an example of a range map produced from the R script. When we put it in the visualization, we will export the data as a MultiPolygon that should be read in fairly easily as either a geoJSON or geoTIFF format.

- Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?
  - We decided to rework our method of input selection. We now only allow users to select input by bird species. We concluded that it would not be practical to aggregate any range, trend, and abundance data by bird type. First, we have no good way to combine separate range and trend overlays. Second, we felt that an aggregation of bird species would oversaturate the visual space, and trends may be harder to find. For example, an aggregate species range view or observation dot density plot may cover the entire continent, even if the individual species habitats are decreasing. Allowing the comparison of only species will make the intended use more obvious and intuitive to users, and will ensure that our plots stay readable and meaningful.
  - We felt that an overlay of various data on a map of North America would be necessary to answer the questions provided in our 'questions' section. Because habitat range and population are tied to spatial location, a map view is an effective visualization choice. We also decided to overlay abundance, range and trend data on this map. We initially decided to show abundance data as a dot-density plot of observations. These observations can be filtered by a year slider to interactively show how observation count changes over time. While this information may show underlying abundance trends, it could be biased by increased observation reporting over the years. Because of this, we felt it was appropriate to display a heat map of expected abundance in addition to the dot-density plot. The heat map takes into consideration the number of observations submitted to provide an accurate estimation of population. We also decided to display range data as an area plotted over our map. We felt this was the best way to display habitat change, as habitat is inherently tied to geospatial position. The last form of data we decided to overlay on our map was trend data. We initially decided to implement the trend map from the Ebird visualization using the P-values as data-points. We decided that this is a non-intuitive way of showing population trend. A high P-value does not always correspond correspond to a large change in abundance magnitude. We thought it would be better to encode the magnitude of change, by showing the percent change in observations from 2006-2018.
  - We also decided to include a beeswarm plot showing the change in population abundance of all species. Unlike our original design, we decided not to expand the bubble chart based on filter category. Because we are only allowing for the selection of individual species, we can simply color the circles corresponding to selected on the beeswarm to show how they compare to the overall trend.

- Last, we decided to include the parallel coordinates plot from our proposal. After looking through the data available to us, we decided to generate statistics to plot on our parallel coords from our abundance data set. For each species, we can precalculate the total number of observations, percent change in observations from 2006-2018, percent change in range from 2006-2018, and bird type. This plot will provide a good overall picture of population and habitat trends. Each axis of the parallel coordinates can be sorted on in order to find species with the most striking trends.

- Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.
  - We have currently implemented an interactive dot-density plot for our observational data. We plan on eventually hosting our data on an sql server, but currently have a small csv stored in our project data folder. We have pulled a small set of observations for a single species to start setting up our map view. Selecting the species from the dropdown will populate the Map with observation points. Each observation record is encoded as a uniform sized circle. The opacity of the circle encodes the observation count of each record. Clicking on a season in the legend will snap the brush to that specified range. Brushing over the season bar will subset the observations based on that season range. The year slider filters observations by year.

- Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?
  - We are still exploring the data to find a solid subset of birds that have interesting migratory patterns. The map and parallel coordinates plots should be able to add the interaction we want in order to make the ebirds plots more compelling. Once completed, our visualization should allow for multiple views of the data that will illustrate how bird abundance has changed over the last decade.