Data Visualization Project Proposal

Will Richards, Corin Thummel

- **Basic Info.**
  - Project Title: Habitat and population trends of North American birds
  - Team Member: Corin Thummel, thummelcorin@gmail.com, u0261133
  - Team Member: Will Richards, richardsw2017@gmail.com ,u0401321
  - Project Repo: https://github.com/cthummel/dataviscourse-pr-birdHabitats

- **Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.
  - 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.

- **Project Objectives.** Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.
  - **Core Objectives**
    - How is the overall North American bird population changing over time?
    - How is the overall North American bird habitat range changing over time?
    - Correlation of loss of habitat with population decrease?
    - How do population trends across different bird types compare?
      - What bird types are most at risk?
    - How do habitat trends across different bird types compare?
  - **Secondary objectives**
    - Any trends of non-seasonal migration?
      - What percent of habitat is conserved?
    - How do population trends across different biomes compare?
      - Which biomes are most/least at risk?
      - Is there a significant difference in population decline across biomes

- **Data.** From where and how are you collecting your data? If appropriate, provide a link to your data sources.
  - Our data will be provided by the Cornell lab of ornithology. The lab has been collecting and curating a comprehensive dataset on bird population, range, location, and migratory patterns for over a decade.
  - We have been granted access to this dataset.
  - https://ebird.org/data/download - You can create an account and request the dataset from here.

Data Visualization Project Proposal

Will Richards, Corin Thummel

- **Data Processing.** Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

The data is hosted on AWS, and can be accessed through prepackaged options, or generated with specific queries. We plan to select a representative sample of species to make data storage more manageable.The main dataset is stored as a tsv with column headers. Many of the data attributes are stored as ids. The dataset contains separate files for the appropriate id-data mappings.

We first plan to use population data for different bird species at geographical points/regions spanning north america. We will need to convert latitude/longitude data to geoJSON. We can use a csv-to-geoJSON generator web application (http://www.convertcsv.com/csv-to-geojson.html) to do this. We will display this population data as a dot density plot.

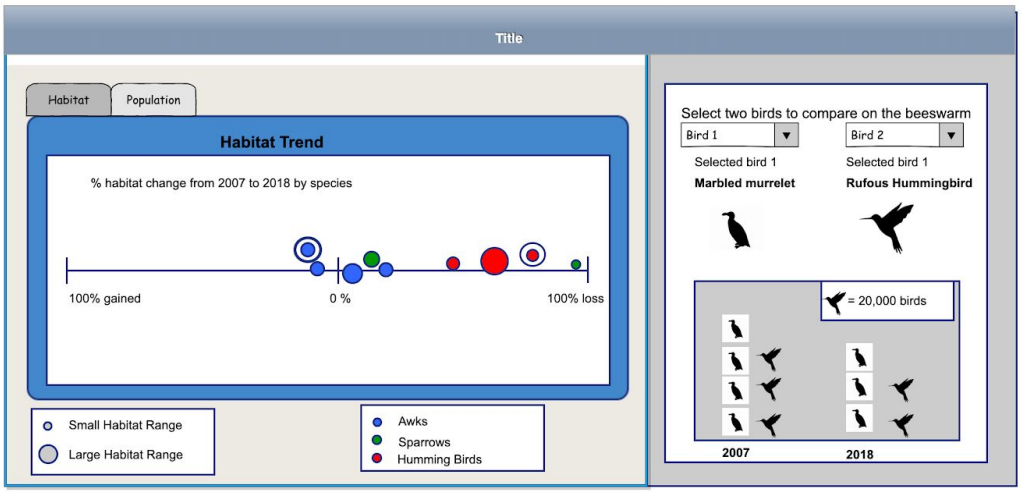
We also need to generate habitat range data from our data set. The range area is composed of 2.8x2.8km grid cells. The range area is available as a geopackage file, which can be converted to geojson using online tools. A species range is an estimate of area occupancy (IUCN). This is the smallest continuous area that can be drawn to enclose all known or inferred citings on the bird.

Last, we need to display trend data. Our dataset provides us with p-values for the significance of population decrease over time. We will directly use these p-values to display trend information.

Data Visualization Project Proposal
Will Richards, Corin Thummel

- **Visualization Design.** How will you display your data? Provide some general ideas that you have for the visualization design. Develop **three alternative prototype designs for your visualization**. Create **one final design that incorporates the best of your three designs**. Describe your designs and justify your choices of visual encodings. We recommend you use the Five Design Sheet Methodology.

- Prototype Visualization #1: Beeswarm and isotype pictogram



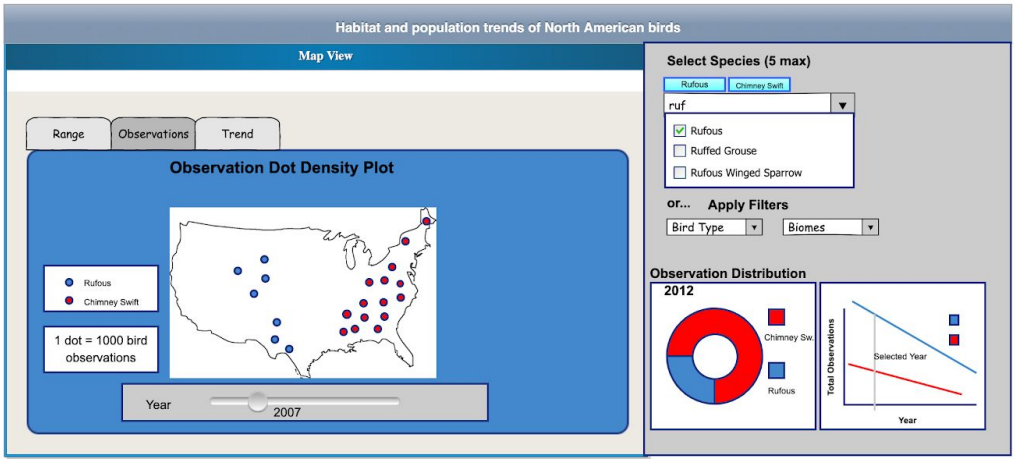
We wanted to create one prototype visualization that did not include a map component. We achieved this by using a bee swarm and isotype pictogram.

The plot on the left is a bee swarm plot with circles encoding different birds from our data set. The area of the circle encodes the size of the habitat range for the bird while the color encodes the type of bird. We use a diverging scale around 0 to show birds that have gained habitat over the last decade versus birds that have lost habitat. There is a tab at the top of the bee swarm plot to change the horizontal variable to show bird population instead.

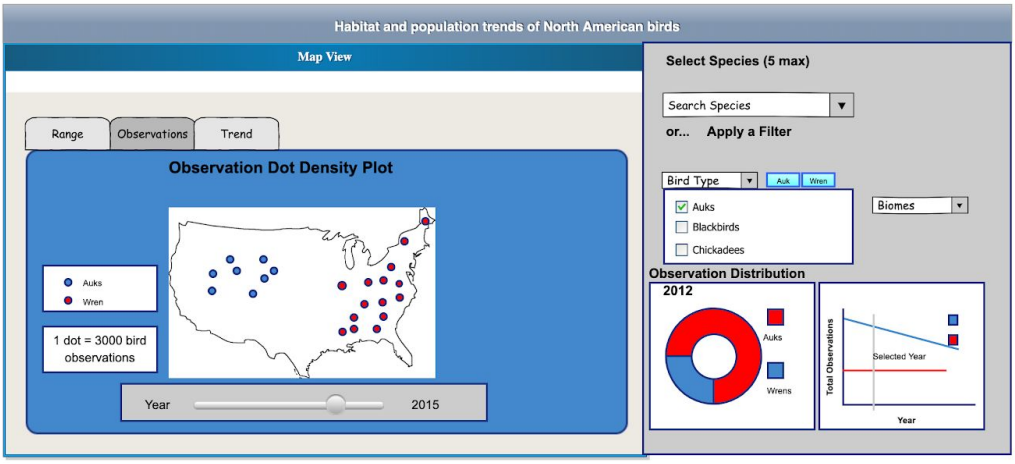
The plot on the right is an isotype pictogram. This plot is a very approachable way to compare the populations of two birds. At the top, there are two drop-down menus to select a pair of birds. When selected, the corresponding circle in the bee swarm is also highlighted. Each bird is encoded as a symbol and we use position as a channel to show relative rounded populations.

Data Visualization Project Proposal
Will Richards, Corin Thummel

- Prototype Visualization #2:Map and scatterplots



In this design, we encode latitude and longitude information by map position. The data plotted on the map will be controlled by one of two input methods. A user can either select which species' to visualize, or they can select data with the 'bird type' or 'biome' filter. A user can search for and select up to 5 species of birds. Any species selected will be displayed as a badge above the selection. If a species is selected, the dots for our data density plot will be generated from species observations clustered by location. Each species type will be encoded with one of 5 colors.

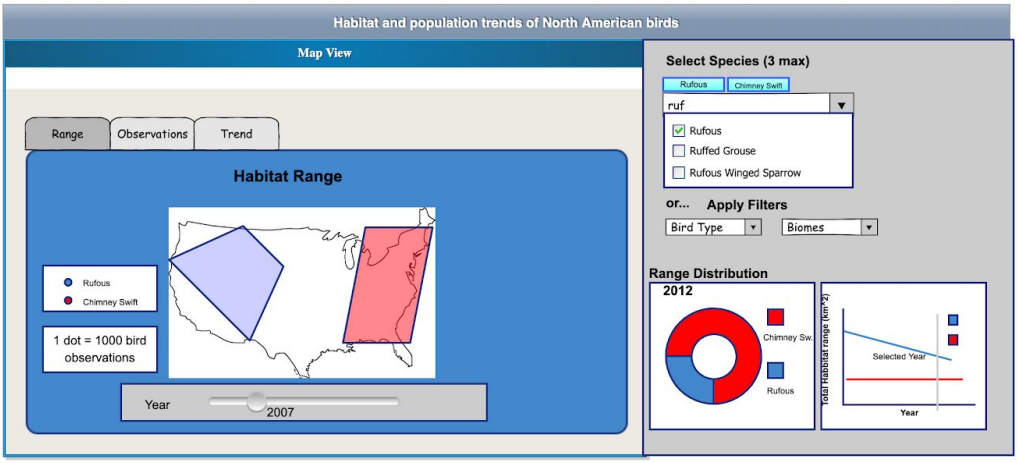


On filter selection, any species and badge selections will be removed. Users can select up to 5 items for a given filter. Any selected filter will be displayed as a badge that is cleared upon

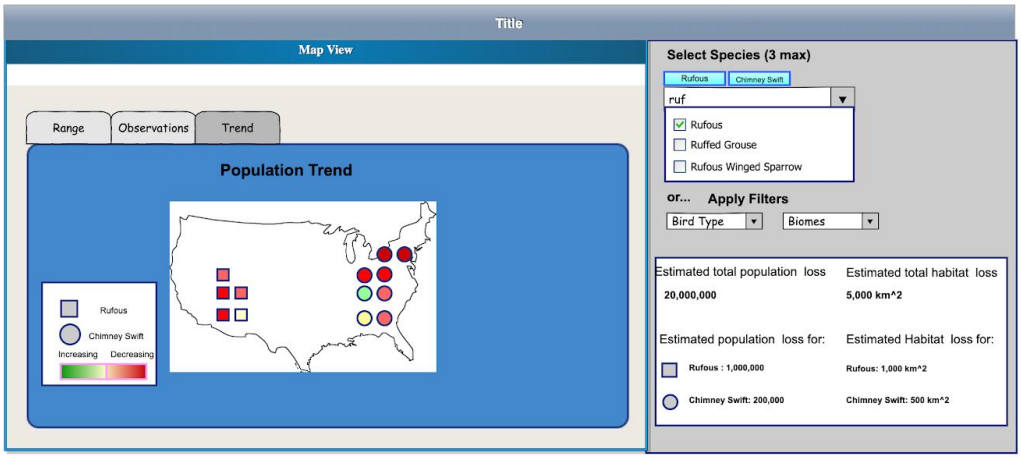
species select or a different filter select. If a filter is selected, dots for our plot will be generated from observations of all birds encompassed by filter type.

If the observation tab is selected, we will overlay an observation dot density plot on of our map. The year slider at the bottom control the display of observations from a specific year. The dot density threshold will be scaled appropriately to different data densities.

If observation is selected, a line chart will display total observations for each user selection (species, bird type, biome from 2007-2018). The linechart will have have a vertical axis showing the selected year. This slice of the line graph is broken down into a donut chart, redundantly encoding the proportion of data each selection contributes at the select year.



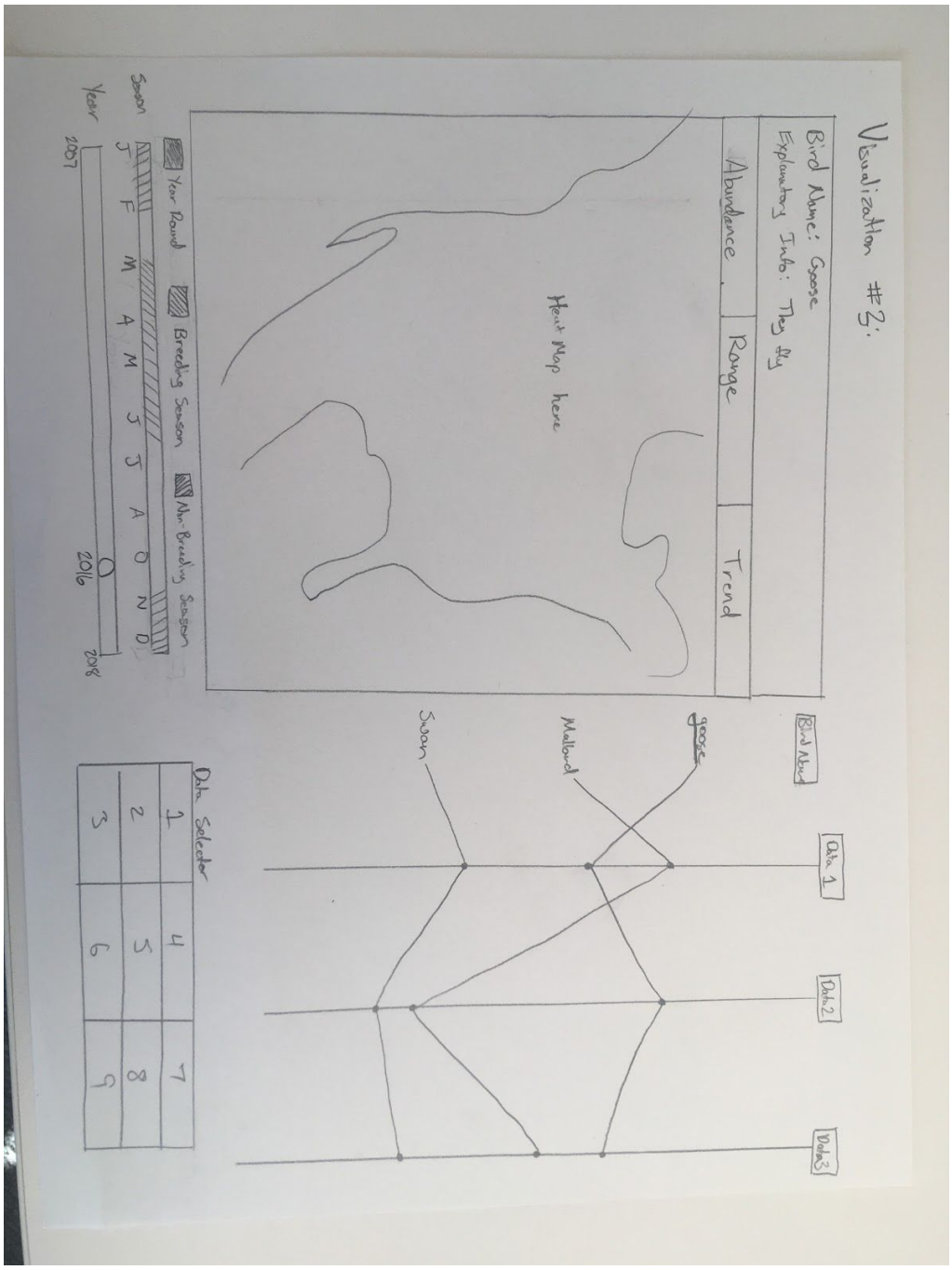
If a user clicks on the range tab, we will overlay areas on the map color coded by input selection. If a user selects species, each species range will be shown with a different color. If a user selects a biomes, the range for all birds of a biome will be displayed with a color. The y axis on the line chart will show total area in km^2 of each selection's range.



The trend tab will display the overall significance of population change from 2007-2018. We use a diverging color scale to indicate if the population is increasing or decreasing. The stronger the significance, the darker the shade on the color scale. We will plot points uniformly across the

selections range. Since we use color to indicate significance, we now must encode selection identity with shape. Since the trend is over a time period, we no longer have need for the year slider. We also decided to replace the donut chart and line chart on the side panel for the trend display. We felt it was fitting to show some overall meta-statistics about the selected data and dataset as a whole. We felt showing the estimated population and area loss for all birds, and for the input selections would help tell the story about habitat and population loss.

- Prototype Visualization #3: Map and Parallel Coordinates Plot:



Data Visualization Project Proposal

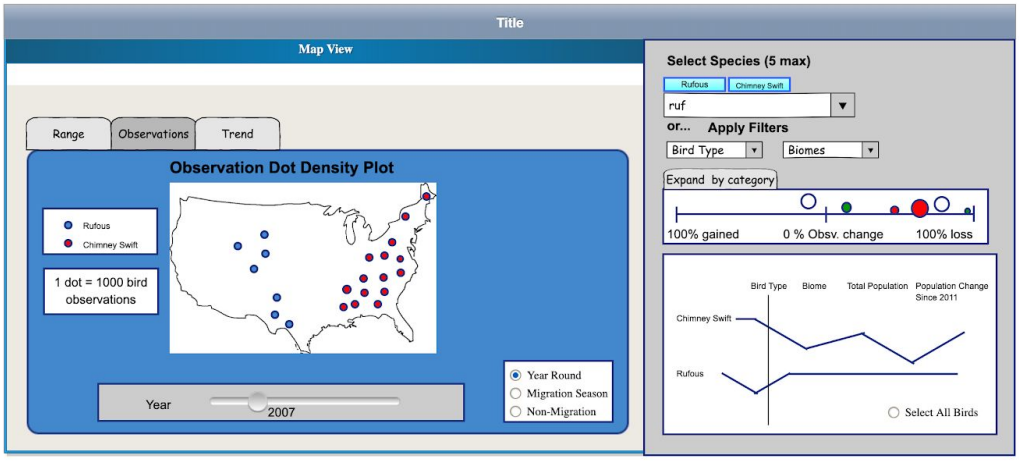
Will Richards, Corin Thummel

- Map:
  - Information box on top for selected bird. Includes name and other information relating to the type of selector on the map.
  - Three buttons corresponding to Abundance, Range, and Trend. Only one of these selectors is active at a time as shown by highlighting the button. It applies to the data shown in the map below with color coding based off seasonal categorical data.
  - The map itself has a heatmap by color corresponding to the type of season (Year Round, Breeding Season, Non-Breeding Season, Pre-Breeding Migration, and Post-Breeding Migration) with a color scale from transparent to opaque corresponding to the data value. Heatmaps are a logical representation for abundance and trend data to show relative density of bird sightings. Since there is a large overlap between various seasons, overlaying multiple colors is a concern. The map will center itself in a region corresponding to the bird but allows for some zooming and scrolling.
  - Below the map is a legend explaining the color scheme used for different seasons. The data is subsetted into Year Round, Breeding Season, Non-Breeding Season, Pre-Breeding Migration, and Post-Breeding Migration. Each of the legend objects is clickable to set the subset to the corresponding season.
  - Below the legend is a brushable month rectangle that allows for subsetting the data from a start date to an end date. The map will update with the proper data spanning that time period. The brushed area subsets the overall data and will update the parallel coordinate display as needed.
  - Finally, at the bottom is a year slider that will allow for updating the entire data set based on which year the user is interested in.
- Parallel Coordinates
  - Include 3-4 coordinates to show values of interest across birds. The data coordinates are selectable from a box of buttons at the bottom to allow the user to select the variables of interest. The left most column contains individual birds grouped by category. Hovering a bird or group highlights its/their path through the parallel coordinate system. Clicking on a bird locks in the highlight as well as updates the map. User can use control + click to highlight multiple birds through the parallel coordinates display (The map only gets updated for the first bird clicked).

Data Visualization Project Proposal

Will Richards, Corin Thummel

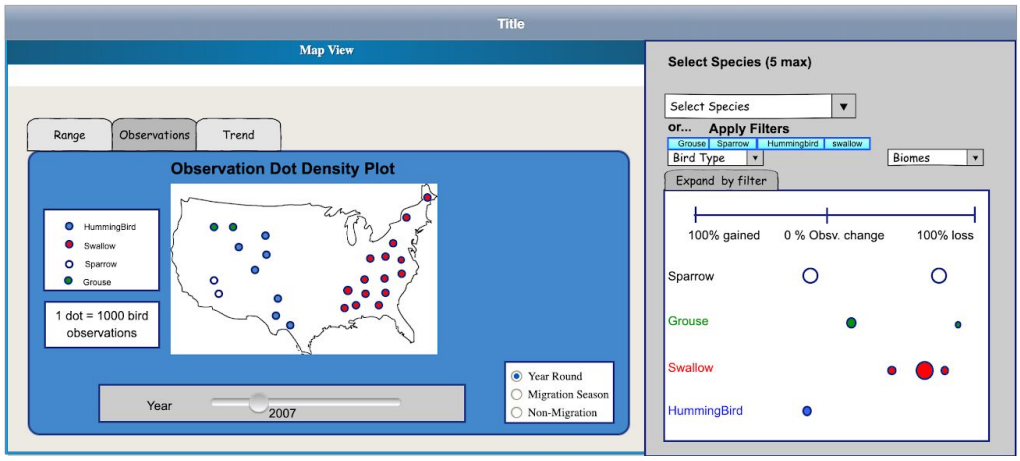
- Final Design:



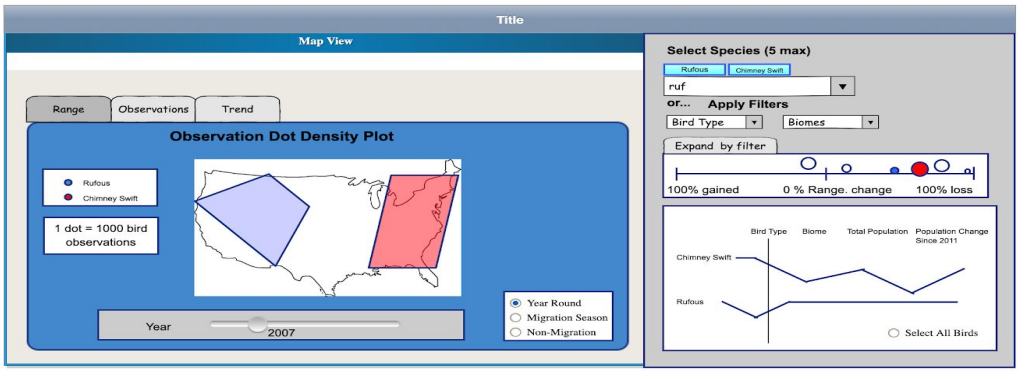
For our final design, we decided to incorporate ideas from all three prototypes. We decided to keep the map view from prototype 1. The design for the map panel has not changed from prototype 1, other than the addition to filter data points temporary based on migration seasons. We felt this would allow the user to notice any changes in migratory patterns that would not be apparent from year round data. Input selection also remains the same from prototype 1. The left hand panel incorporates ideas from prototype 2 and 3. The beeswarm plot encodes the total change of a selected value from 2007-2018. If the observation tab is selected, the beeswarm plot encodes the total change in selection observations over this time frame. If species was chosen as input, we will display a beeswarm plot and a parallel coordinates plot for the species. The beeswarm plot will be unable to expand by filter category. Each selected species will correspond to a single point on the beeswarm, and will be encoded by color corresponding to the map legend. Any species that is not selected will be white. Our parallel coordinates plot will have our input selections as its row headers. The columns will encode values for ‘bird type’, biome, total observation change over time, total habitat area change over time, percent population change over time, and percent habitat area change over time.

Data Visualization Project Proposal

Will Richards, Corin Thummel



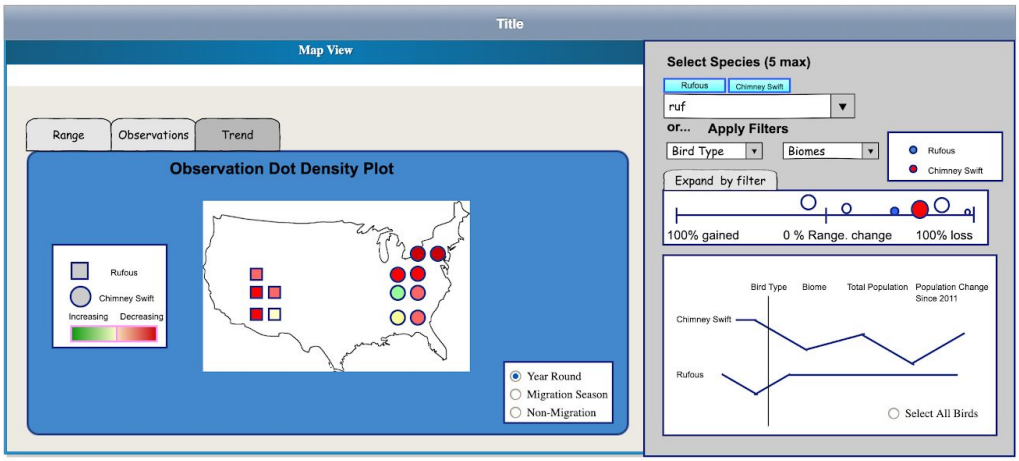
If a filter is used for selection input, we will display a collapsed beeswarm and parralel coordinates plot, but provide the option to expand the beeswarm and replace the parallel coordinates plot. If a filter is selected, then we can separate out the beeswarm by filter type. This will allow us to easily compare overall trends between our selected filters.



If range is selected, the x-axis on the b-swarm will correspond to the percent of habitat area lost or gained for each selection.

Data Visualization Project Proposal

Will Richards, Corin Thummel



If trend is selected, we will show the p-value of population change over time ("Note the axis on the figure above is incorrect). A high p-value for population increase will be on the left, and a high p-value for population decrease will be on the right.

- **Must-Have Features.** List the features without which you would consider your project to be a failure.
  - Map
    - Abundance, Range, Trend overlays.
    - Some zoom capability
  - Year Slider for subsetting data by year
  - Seasonal Brush for subsetting data by season
  - Beeswarm plot colored by bird category with position based off map overlay data and circle size based off interesting data. Plot must expand to show individual bee swarms for each bird category.
  - Parallel Coordinates plot with three coordinates that are by default the same coordinates as the condensed beeswarm plot.
  - Users must be able to subset the data by selecting a single bird. This updates the map to show data corresponding to the bird and the birds values are highlighted in the beeswarm and parallel coordinates plot.
  - Users must be able to subset the data by selecting multiple birds. This can be done by brushing in a beeswarm plot or using control + click on points in the beeswarm. The map and parallel coordinate plots will then update to display data from each bird.
- **Optional Features.** List the features which you consider to be nice to have, but not critical.
  - Storytelling aspect. We could walk the user through a particular example that shows each button to click and what conclusions might be drawn from the data. This could be a multiple step story.
  - The Beeswarm and parallel coordinates plots have customizable data fields. In other words, the user can choose data types that sound interesting to them and use that to resize the circles in the beeswarm, change the horizontal position, or maybe the

color scheme. Similarly in the parallel coordinates plot, the column categories could be changed to fit the user selections.
    - Map allows satellite, street or hybrid views allowing users to see the exact location of each bird sighting.
    - Integrate with ebirds api to show real-time observations. This could be shown on the map by drawing a circle at the location with a mouseover tooltip that explains in detail what the new entry was. We could set the circles to disappear over time or only show new sightings of selected birds. We would have to be careful not to overload their API since they somewhat limit the number of requests.

- **Project Schedule.** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.
    - Nov 1st: Have data processed and rolled up into useable JSON objects.
        - Subset data by hand - WIll
        - Process data - Corin
    - Nov 8th: Have initial map view complete with trend, abundance, and range overlays.
        - Range map - Will
        - Abundance map - Corin
        - Trend map - Will
    - Nov 15th: Build side panel
        - Beeswarm - Corin
        - Parallel Coordinates - Will
    - Nov 22nd: Develop filters, and storytelling aspect
        - Biome and bird type filters - Will
        - Storytelling - Corin
    - Nov 27th: Finalize all features and polish
        - All hands on deck