European Birds – Information Visualization

Nico Koltermann
Lisbon, Portugal
nico.koltermann@tu-dortmund.de

Emma Knierien
Lisbon, Portugal
emma.dennisknieriem@gmail

Matheus Trindade
Lisbon, Portugal
matheus.tri.2010@gmail.com

INTRODUCTION

This project addresses the properties of European bird species that are most influential on their current classification of extinction risk, according to the International Union for Conservation of Nature (IUCN). Understanding factors such as their natural environment, countries, and whether they are hunted can be valuable assets in the fight to preserve species against extinction. This project aims to analyze relations between population size and surface areas, the risk of extinction over time, which bird families were associated with a greater risk, and mapping the risk of extinction on the European map. This information, though published in a public dataset from the European Data Agency - Article 12, 2020 Dataset Reporting, is no synthesized well enough in the literature in order to provide clear answers to stop extinction.

Some questions related to this domain were initially listed, as follows:

1. Are only individual populations threatened or also entire families? If so, which families?
2. Within threatened populations, what is the proportion of species with large surface areas?
3. Out of the wintering species whose individual populations are less than the first quartile of all wintering birds, what proportion has a risk of Endangered (EN) or higher?
4. Which bird populations have the biggest differences in their total population size during the research time? Analyzing the top 50 variations, can it be a trend of extinction risk?
5. What's the average of the species' population that are currently classified as threatened according to the red list category (CR, EN, VU)? If a bird watcher picks a random specie to study, what's the probability that the population of this specie will be above the calculated average?
6. Analyzing the proportion between species that are nationally hunted and those that do not belong to the LC (Least Concern) Group, can it be pointed as a trigger for birds extinction too? (merging data between tables).
7. Which is the most hunted family according to the dataset (merging data between tables)? Which country should focus more efforts on preservation according to the number of families under extinction risk?
8. Does there seem to be a correlation between species' populations and their "surface areas"?
9. Is it possible to confirm a threatening trend, by comparing both attributes "red_list_cat" and "red_list_cat_prev?

RELATED WORK

The data used in the project is taken from the European Data Agency. The endangerment threat level from this source was taken from "IUCN Red List categories and criteria, version 3.1, second edition", compiled by the IUCN's Species Survival Commission (SSC). During the research phase, it was found other projects that focused not only on European bird species' endangerment, but that based their work on the same IUCN data.

BirdLife International is an ornithological conservation group that authored the report "European Red List of Birds 2021" using the IUCN data. There are two pie charts in Figure 1, one for the continent of Europe and one for the European Union, since the EU was a contributing organization to this report.[6]

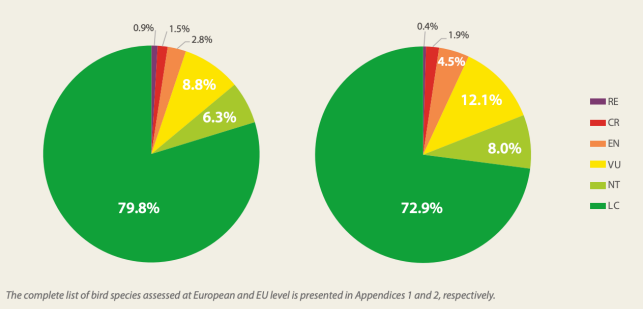

Figure 1. Bird species threatening classification according to BirdLife International, 2021. In the left it's being represented the European continent and in the right, the European union.

In each figure, the two comparative pie charts are largely the same because the EU includes most of Europe; 28 countries at the time of the publication. The information the pie charts express corroborate the findings a viewer can draw using the visualizations from this project, but to a much lesser degree of detail. These graphics fall short in the amount of information they can convey because each figure encodes only 2 attributes (percentage and category to area and color hue, respectively). All of the other visualizations in this report are similar, with the exception that some encode counts instead of percentages.

Tyhe Reading's "Threatened Species Across the Globe" visualization domain is similar to the one studied in this paper and BirdLife International's but includes global data, not just data from Europe.[7]

Figure 2. Bird species threatening classification according to Tyhe Reading's "Threatened Species Across the Globe".

While Reading's vis encodes more information than BirdLife's, it still fails to communicate efficiently and instead suggests relationships that are inaccurate. Every country included in this vis is in a circle radially, but it is unclear how they are organized, other than being grouped by continent. Every mark is connected to the mark of the same threat level of the countries on either side of it with a line segment, but there is no geographic or other apparent relationship between adjacent countries. Counter-intuitively, the radial scale (presumably representing the number of species, though this is unclear) decreases from the innermost to outermost concentric circle.

"Threatened Species - Birds 2020 Animated" by Sandraviz cites OECD (The Organisation for Economic Co-operation and Development) data, which itself references the Red List Categories from the IUCN. [5]

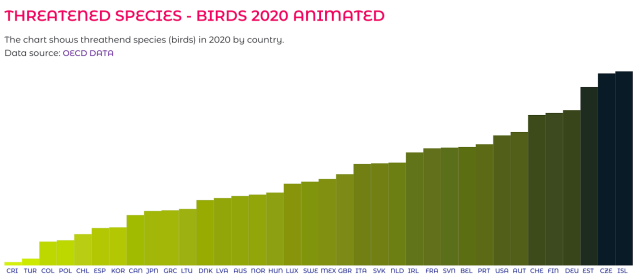

Figure 3. Threatened bird species per country according to The Organisation for Economic Co-operation and Development.

This visualization uses the D3.js library, but like the aforementioned two sources, is simplistic and offers the viewer little insight. As the bar graph's counts increase by country from left to right, the color gradient darkens proportionally, suggesting that this is also a case of double encoding (though again, this is unclear). Unlike the other two, this visualization is interactive. The bar graph reanimates upon loading the page or adjusting the window. This limited user interaction is a purely aesthetic touch and does not encode any information in itself.

The visualizations whose domains relate to this work can give basic information about endangered bird species' counts, but are very simple and can also be misleading.

THE DATA

The data consists of two datasets that were obtained from European Data Agency - Article 12, 2020 [1], named as "Article12_2020_birdsEUpopulation" (main dataset) and "Article12_2020_data_birds_annexII" (secondary dataset). The data was cleaned according to the following methods:

1. It was used the sentinel method to treat missing values in the "Distribution surface area", using filter in excel.
2. The attribute "red_list_cat" represents the International Union for Conservation of Nature Red List of Threatened Species Category, that consists of byssilabic terms such as LC, NT and VU. The data had a few items in which the author wrote observations beside the category name. These observations were not useful for this study (example: VU A2bcde+3bcde+4bcde), and they were cleaned using excel filters (outcome: VU).
3. 18 attributes in the main dataset were deleted using Python. These attributes presented data not needed to answer the questions related to the domain.
4. It was used Python to merge a second dataset to the main one, in order to assimilate information regarding hunted species and species per country. Even though the information was merged through the attribute "speciesname", it was observed that the second dataset didn't have all the species listed in the first dataset, and due to this, after merging, the information regarding many species was lost. It was implemented the strategy of applying the main dataset(non-merged) to the Dot Matrix, Scatterplot, and Sankey Diagram, showing 100% of the original items, and the merged dataset to the Choropleth Map, showing the countries and through interactivity, the items associated to the respective countries.
5. Further outliers were analyzed using excel filters, it was identified very small numbers for Population Size (example 1 and 5), and Surface Area (example 100 and 200), but in all cases it matched a high extinction risk, and a very low population number also indicates a very low area.
6. During the application of the Scatterplot filters, it was observed that some items from the attribute "red_cat_list" had a space after the last letter, and that was causing issues with the graphic filtering. So it was required one more

cleaning through excel, replacing, for example, all "CR " for "CR".

7. For the item with attribute "speciescode": A413, the attribute "red_list_cat_prev" was not evaluated in the preview study, being the only information missing. It was applied the senthinel method to treat it.
8. The items with attribute "speciescode": A460-X and A044-X had no evaluation for "red_list_cat" and "red_list_cat_prev" in both studies. Since the main goal of this project is to evaluate characteristics of the threatening risk related to birds' species, the lack of this information made these items not valuable for the visualization purpose, and they were deleted.

During the design of the Sankey Diagram, the main idea was to analyze how many and which species were under threatening risk across time. The main dataset presented information regarding the study of 2020 (red_list_cat) and 2012 (red_list_cat_prev), but the goal was to merge datasets from other years, like 2015, and have a more complete data analysis. However, the dataset from 2015 didn't apply the IUCN denomination as a Threatening Risk Category Scale, and after many efforts of converting it, it was not possible to merge the data.

Initially it was conceived an idea of a derived variable that would classify the birds' covered surface areas as large or small, and allow the user to generate conclusions from that. As the prototype designed happened, it was observed that a scatterplot would communicate this idea in a more effective way, that led to a change in question 2.

During the design of the scatterplot, it was planned to implement a Cleveland Dot Plot to represent the minimum and maximum size of species population during the study. However, after the implementation of the scatterplot, it was observed that distribution was already completely crowded with dots. Besides that, the graphic was already encoding three attributes, with buttons and the quartile information, and any further attribute would probably compromise the understanding of this idiom. Therefore, it was selected only the attribute "population maximum size" to work with, representing the best-case scenario for the European bird populations during the study, and the attribute "population_min_size" was discarded.

After the filter for the attribute "keywintering" was implemented in the graphics, it was noticed that all the wintering bird species, that represent the bird species that migrate during winter, were not shown in the scatterplot. After analysis of the data, it was concluded that the reason they had no items representation was because the natural "Surface_Area" couldn't be estimated due to the migration process, and therefore, all the values for this attribute were sentinel values treated during the data pre-processing.

VISUALIZATION

Overall Description & Rationale

The visualization includes 4 boxes for the idioms: a Choropleth Map, a Sankey Diagram, a Scatterplot, and a Dot Matrix. On the top of the page there is an overall legend the IUCN categories studied, and the hue associated to them in the project. This is displayed only for the user's information.

Through interactivity, the user can see visually encoded information related to the attributes: "Keywintering" (regarding winter migration), "taxFamily_en" (birds' families scientific name in English), "Surface_Area"(area covered by the bird species), "Population_max_size" (maximum population size of a species during the study), Red_list_cat (IUCN category), "Speciesname", and "Country"(Countries where the species are found). When the user clicks on the Scatterplot IUCN buttons, only the data related to this category is colored and the further items are turned grey, and the same logic applies for the "Keywintering" button. Aditionally, hovering over the dots and countries will provide tooltips with important information about the items, and also highlights the item that is being hovered.

The "red_list_cat" attribute is encoded in the hue of the dots and lines, in the graphics: Scatterplot, Dot Matrix and Sankey Diagram, this is the center of the study and connects everything. It was applied a diverging color scheme between green and red, automatically giving the user the notion of what is "alarming" as red, and what is "safe" as green.

The attributes "Population_max_size" and "Surface_Area" are encoded in the Y-axis and X-axis of the Scatterplot, respectively. It was implemented interactive quartile lines for the attribute "Population_max_size", that shows the quartile of the group that is selected through the IUCN buttons. (Example: the quartiles of the populations of bird species that are vulnerable). The encoding strategy for the axis helps with a clear visualization of the relationship between area and population of the birds.

In order to optimize the scalability, it was implemented a log in both Scatterplot X-axis and Y-axis. Initially, the distribution of the dots was either not fitting in the graphic space or unreadable, being all the dots over each other.

The attribute "red_lis_cat_prev" was encoded in the first column of the Sankey Diagram, and the attribute "red_list_cat" in the second column. The items are encoded in lines.

The Choropleth Map was designed to show the number of threatened species reported by the EU countries (Attribute "red_list_cat", categories VU-Vulnerable, CR-Critically Endangered and EN-Endangered), per country in Europe. The sum of the items related to these 3 properties of this attribute, is encoded in the color saturation of these countries (Hue: Blue, varying from light to dark according to scale. This encoding strategy aims to provide immediate information to the user regarding the countries that require more efforts in the birds' preservation.

During the development of the layout, it was initially implemented a background with a forest image, illustrating the context of the project. But as all the graphics were created, the background became chartjunk, calling more attention than the content of the graphics, and it was removed.

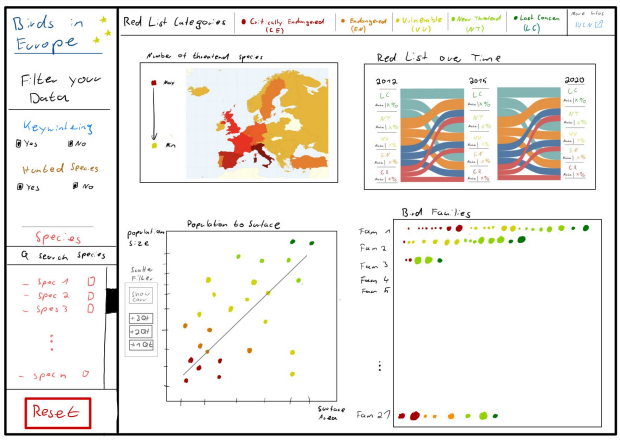

Figure 4. Initial Prototype

In the initial prototype, it was planned to implement two additional filters: a searching filter for "speciesname" and a filter "YES/NO" for Hunted species. Regarding the searching filter, it was found challenges in the communication between this filter and the Scatterplot buttons plus the "keywintering" filter, that put in risk the good functionality of the prototype. Due to this reason, this idea was changed, and the filter removed.

The attribute "nationally hunted" was merged from the second dataset. As mentioned before, the merging process between the datasets led to a loss of many items, since the secondary dataset didn't have all the information about the species present in the first one, since this information was all reported by the countries of the EU. Therefore, the implementation of this filter could lead to wrong information regarding the birds' hunting status, as the user would select the option "Not Hunted" and see all the not hunted species plus the species with no information reported. This idea was discarded and the questions 6 and 7 updated. The merged dataset was enough to implement the choropleth map, but as can be seen by selecting the "CR" IUCN button, no critically endangered species were reported by any country, and that's a research opportunity for future works.

Custom Visualization

The attribute "taxFamily_en" was encoded in the vertical axis of the Dot Matrix. The attribute "Keywintering" was encoded in the shape of the dots (stars for "Yes", circles for "No") and this encoding strategy was chosen to avoid confusion with the colors of the "red_list_cat" attribute and still providing an immediate perception to the user about the wintering species. The items were encoded in dots, and organized in lines, according to the birds' families, and this encoding strategy helps to understand the threatening risk also in a bird's family level. The tooltips encoded the attributes: "Species_code"; "Speciesname"; taxFamily_en"; "Surface_area"; "Population_max_size" and "Red_list_cat". The tooltips were only applied to the dotmatrix and not to the scatterplot since the dots are more organized in the first idiom, while in the scatterplot their distribution is a bit agglomerated and could be harder for the user to identify the origin of it.

During the development of the "Keywintering" attribute filter, the shapes were changed many times until reaching the star configuration, as the main goal was to make it as different as possible for the user perception, since the size of it, that was also increased to the maximum possible, was also a challenge.

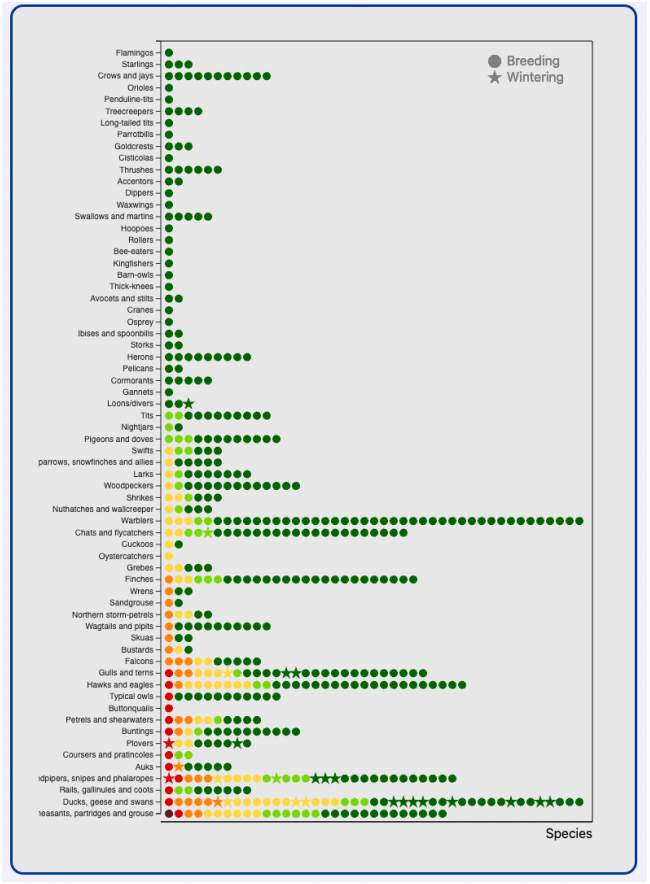

Figure 5. Custom Visualization: Dot Matrix.

Demonstrate the Potential

The development of a few case studies was implemented as a way of testing the potential of the visualization to answer the main questions from the domain.

Question 1: Are only individual populations threatened or also entire families? If so, which families?

The preattentive visual propertie of the dots' hues in the Dot Matrix helps the user to immediately identify the threatened species, by following the legend in the header. The encoding of the attribute "taxFamily_en" in the vertical axis provides

an immediate analysis of the threatening risk per family, where proportions can be quickly and easily estimated. It's possible to conclude that there's one entire family threatened, the *Oystercatchers*, that has only one item associated to it, and this species is currently vulnerable. As a second possible observation, proportionally, the second family with higher risk associated to it are the *Bustards*, with 66% of its members at risk (one species is vulnerable and the other endangered). Another interesting statistic withdrew from the graphic is that the family with higher number of species threatened are the *Ducks, geese and swans*, that possesses 17 members under risk.

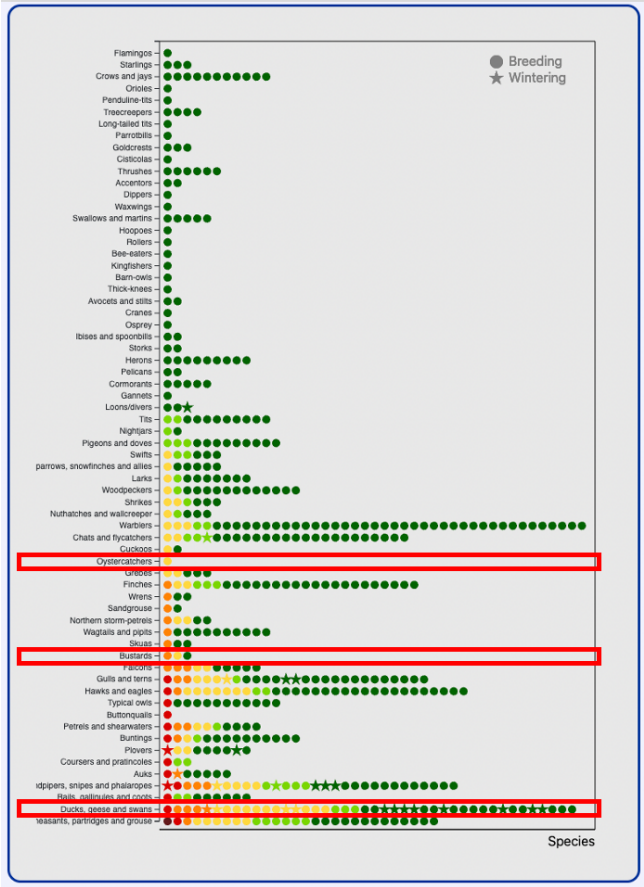

Figure 6. Birds Families' analysis in the Dot Matrix

Question 3 (Updated): A species is classified as endangered when its population has declined between 50 and 70 percent. This decline is measured over 10 years or three generations of the species, whichever is longer, and from this threatening level or higher, the probability of extinction within the next generations becomes alarming for biologists [4]. Analyzing the non-wintering birds, that have a clear relation: "population per area" in the scatterplot, what's the proportion of endangered (oh higher risk) populations within each quartile interval? What does it say about the populations present in the dataset?

The interactivity present in the button "Breeding" in the Dot Matrix allows the user to select only non-wintering birds in both Scatterplot and Dot Matrix, while the other items hues are turned grey. As next, the user can access only the items within the category Endangered and Critically Endangered, by clicking on the respective buttons in the Scatterplot (the desired buttons should be colored), and both interactivities work together, to color as grey items that are not a match for both selections at the same time. Through this analysis, it's possible to observe that from the 483 non-wintering species:

- Under the first quartile: 23,18% of species population are endangered of critically endangered.
- Between the first and the second quartile: 2,48% of species population are endangered of critically endangered.
- Between the second and third quartile: 3,31% of species population are endangered of critically endangered.
- Above the third quartile: no endangered or critically endangered populations identified.

Through this analysis it's possible to observe that the 25% of the items with higher population values present no endangerment risk. Only 7,24% of the total items analyzed and under endangerment risk or higher. 23,18% of the 25% smaller populations are impacted.

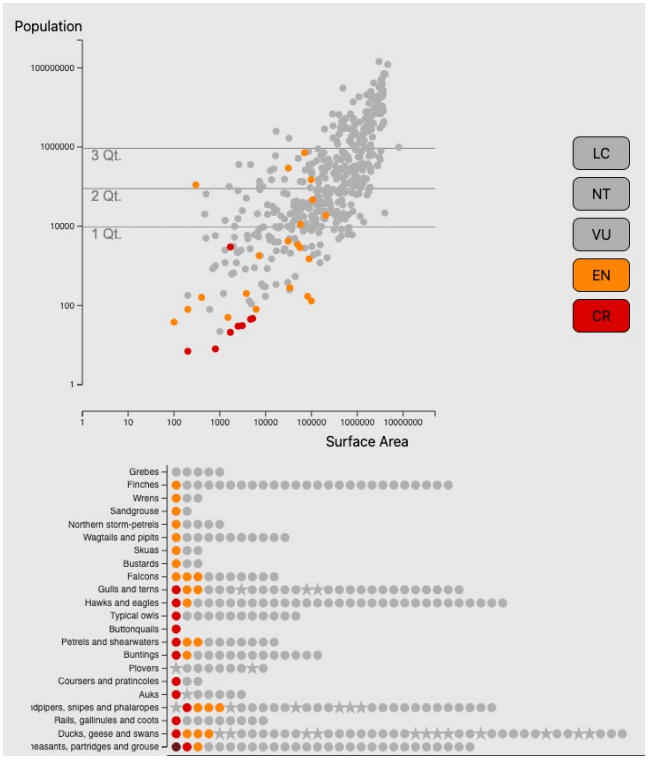

Figure 7. Analysis of Endangered and Critically endangerment non-wintering species (Scatterplot and Dot Matrix)

Question 4 (Updated): Which are the EU countries that need to focus more efforts on birds' preservation? How many threatened species do they have? Can you tell how this threatening risk is characterized by country, for example, does Germany have more vulnerable or endangered birds?

The choropleth map provides preattentive information regarding the countries with a highest number of threatened birds, encoded in the color saturation of each country. Through this configuration, it's possible to tell that the EU countries that need to focus more efforts on birds' preservation are Sweden, Denmark, UK and Belgium, since they have more than 16 threatened species each. By hovering over the map, a tooltip is shown with the name of the country, and the amount of threatened species within the country. Besides that, the country is highlighted in a lighter color and also all the threatened species that belong to this country are highlited in the Dot Matrix and the choropleth map. Germany has 11 threatened bird species, from that 18,2% are endangered and 81,8% are vulnerable.

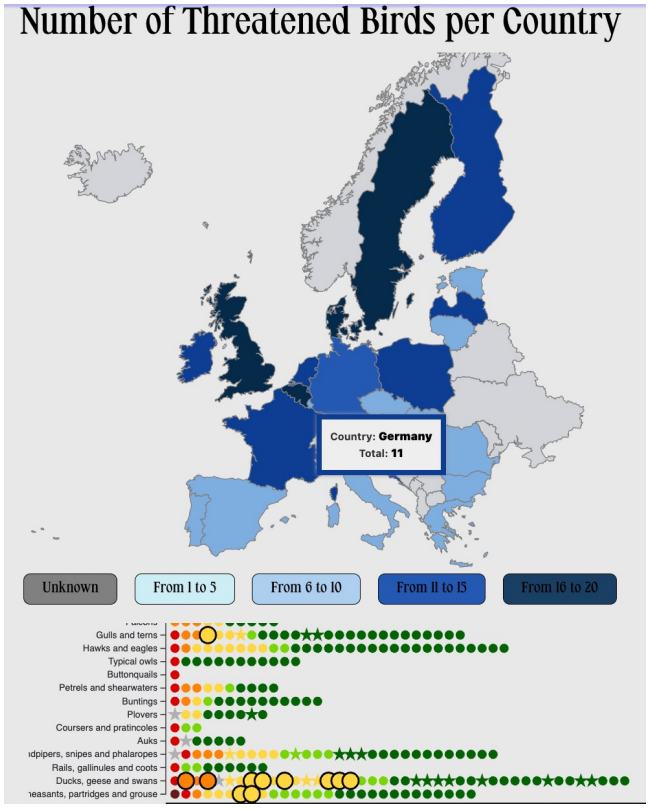

Figure 8. Scatterplot: Analysis of Endangered and Critically endangerment non-wintering species

Question 7 (Updated): Comparing the data analyzed in question 3, is it possible to affirm that birds that migrate during winter have a natural advantage in relation to those that don't migrate when it comes to endangerment risk?

Applying the same strategy of question 3, but selecting the button "wintering" this time, it's possible to generate a statistic about wintering birds' endangerment, corresponding to 15,38% of their total population (analysis of channels: hue and shape), while only 4,5% of non-wintering birds present endangerment risk. Through this analysis it's possible to infer that wintering birds are associated to a higher risk.

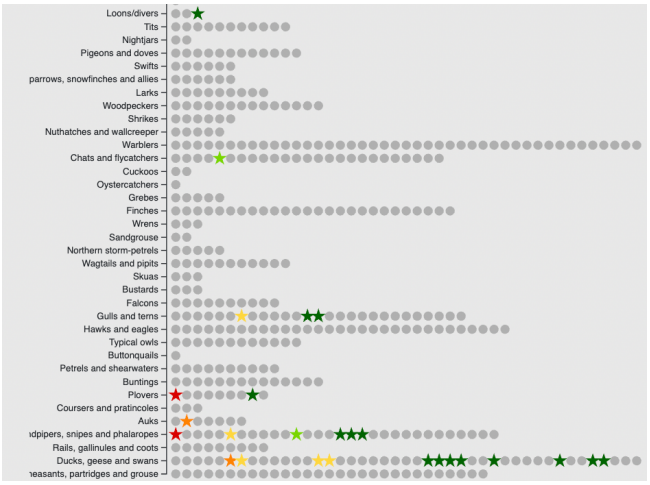

Figure 9. Analysis of risk associated to wintering birds

Question 8: Does there seem to be a correlation between species' populations and their "surface areas"?

The log of the attributes "Population_max"size"and "Surface_area" are encoded in the Y-axis and X-axis of the Scatterplot respectivelly. This encoding strategy gives an immediate perspective of the relationship of both attributes. By looking at the Scatterplot it's possible to observe a clear linearity in the distribution of the items.

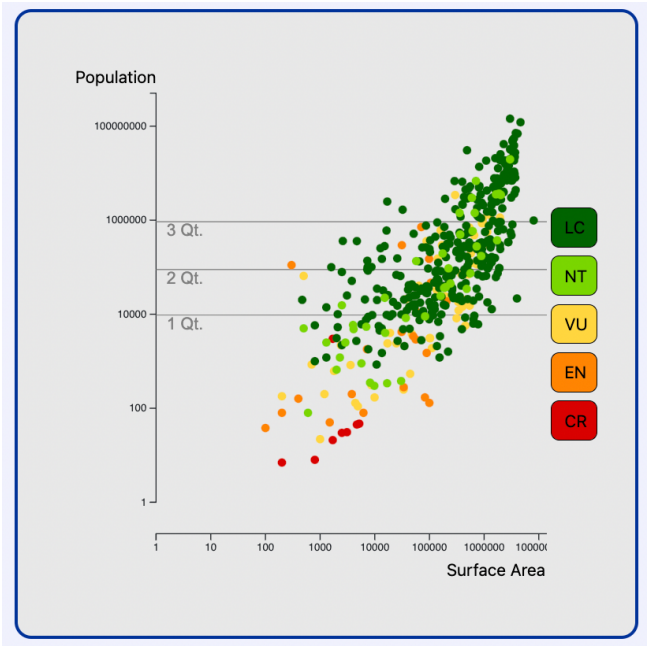

Figure 10. Correlation between species' populations and their "surface areas through Scatterplot analysis

Question 9: Can we analyze the threatening trend, by comparing both attributes "red_list_cat" and "red_list_cat_prev"(representing the studies of 2020 and 2012)? Which species had worse classifications? Which one improved?

In the Sankey diagram, the items are encoded in the lines, and it's possible to have an overall view of how the threatening classification of the birds' species evolved through time. It's possible to conclude that in 2020 80% of the species are out of threatening risk (LC+NT), against 74% in 2012, a great improvement of 6%. The main change that can be seeing is the reduction in the number of endangered species, from 11% in the preview research to 5% in the actual one. Besides that, by hovering over the dots in the Dot Matrix or Scatterplot, it highlights both the dot and the respective item's line in the Sankey Diagram, showing the information of how better or worse this species is doing related to the last research. For example, it's possible to conclude that the species code "A461" improved from the category "Endangered" to "Least Concern", while the species code "A115" had a decrease from "Least Concern" to "Endangered".

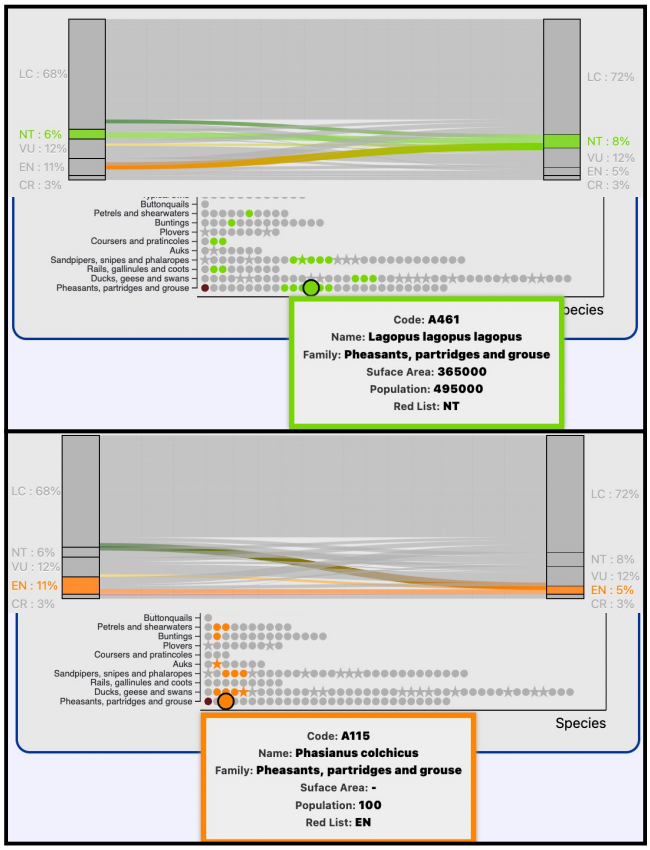

Figure 11. Sankey Diagram: The threatening trend of species (2012-2020)

IMPLEMENTATION DETAILS

The most challenging part of the implementation was handling the different filters. A state machine was introduced for this, to save every filter state ("red_list_cat", "keywintering", "breeding-non wintering"). Depending on its current state, all idioms are updated and the color of the items is selected. They are turned grey if they are not selected, or the respective "red_list_cat" color if they are selected. Based on these changes, the data of the quartiles in the Scatterplot and the map gradient are also updated. It was important to have one main point to handle the data and then afterwards call the different update functions, so possible inconsistencies can be avoided.

The hovering behavior is handled separately. It's applied a function, that handles the view manipulation for all idioms. If the user hovers over a point on the scatterplot or the DotMatrix, the same function just shows data on the whole dashboard.

During the implementation of the Scatterplot, it was initially challenging to fit all the items in the graphic, because they presented a very condensed distribution, and some points would also be too spread. In order to solve that, it was applied a log to both attribute values encoded in the axis, that provided a distribution way easier to analyze.

The values encoded in the color saturation of each country in the choropleth map are not items from the dataset, differently from the Scatterplot and Dot Matrix, but sums of items from the attribute "red_cat_list", filtering off the categories "LC" and "NT", and associating that to the attribute "country". To make this possible, it was necessary to create two arrays "country_summary" and "country_count" to compute these values.

For a better layout, a config file handle all layout information for each idiom, as svg height/widht, colors, radius of points in normal state and hover state. With this config file, it was easier to adjust the layout without searching the whole code.

For the choroplath map and the sankey diagram, the d3 library was extended by these extension, compatible to the d3v7 used in this project.

In order to present some styling in the dashboard, the bootstrap library was included. Then it was possible all the pre-designed classes for layout like the flex-classes or the styles of the buttons.

The project was also structured in many different files. Each idiom was developed in its own file, the interactions and updated function are in different files. Also helper classes were implemented to handle generic task, that are useful over the full project.

CONCLUSION

Based on these findings, most bird species in Europe are of little concern regarding endangerment and extinction. The interactive data visualization based on IUCN data suggests that overall bird species endangerment health has declined from 2012 to 2020, species population increases as its surface area increases, and that bordering countries often have similar numbers of threatened species.

While the majority of species fall into the two least threatened Red List categories, dozens of species and millions of individual birds are at risk. This affects not only avian populations but has potentially severe consequences for remaining European conservation areas, as everything in

nature is part of complex and dynamic systems and subsystems.

The visualization is able to answer most of the original questions. Some, though, had to be modified or replaced. Adjusting our expectations is a normal part of learning more about, exploring, and making inferences from a dataset. It was the right decision to replace some of the questions since the first checkpoint, as it was spent more time thinking about the relationships between attributes and formed more well-informed queries.

One thing that should be done differently in future projects is to find more comprehensive data about hunting. It was found challenges merging the main dataset about countries, species, and threat level with the hunting data. Several rows from each original dataset were incompatible with the other dataset, necessitating the omission of these rows in order to keep the merged dataset consistent.

Further time and funding would be spent on incorporating more potential threat factors into our data to determine which have the largest impact on endangerment to European bird species. In addition to hunting information, it's believed that more data about air quality/pollution and deforestation/industry building would be particularly relevant to bird population size, and could be more insightful about endangerment than hunting alone [2]. In order to avoid our previous issues of merging data, commission conservation experts could conduct pollution/deforestation studies, designing each study with the goal of adding to our existing IUCN threat level data for a smooth merge. Air pollution and deforestation studies have already been done, such as "Aerosolized Coal Fly Ash: A Previously Unrecognized Primary Factor in the Catastrophic Global Demise of Bird Populations and Species" and "Using remote sensing to inform conservation status assessment: Estimates of recent deforestation rates on New Britain and the impacts upon endemic birds" [3]. In order to incorporate new information to the dataset(s), it's needed information specific to Europe and carefully aligned with the existing data so that it's possible to include all information without omission.

The visualization already includes one species that has now gone completely extinct, and if more follow the declining trends, we can only expect to lose more species to history.

REFERENCES

1. European Environment Agency. Status and trends of bird populations, 2020. Retrieved September 23, 2022 from https://www.eea.europa.eu/data-and-maps/data/article-12-database-birds-directive-2009-147-ec-1.
2. Graeme M.Buchanan, Stuart H.M.Butchart, GuyDutson, John D.Pilgrim, Marc K.Steininger, K. DavidBishop, PhilippeMayaux. Using remote sensing to inform conservation status assessment: Estimates of recent deforestation rates on New Britain and the impacts upon endemic birds. Issued January 1, 2008. https://www.sciencedirect.com/science/article/abs/pii/S0006320707003564
3. Mark Whiteside, J. Marvin Herndon. Aerosolized Coal Fly Ash: A Previously Unrecognized Primary Factor in the Catastrophic Global Demise of Bird Populations and Species, 2018. Published November 10, 2018. https://www.researchgate.net/profile/J-Herndon/publication/328859493_Aerosolized_Coal_Fly_Ash_A_Previously_Unrecognized_Primary_Factor_in_the_Catastrophic_Global_Demise_of_Bird_Populations_and_Species/links/5ec868fca6fdcc90d68fa992/Aerosolized-Coal-Fly-Ash-A-Previously-Unrecognized-Primary-Factor-in-the-Catastrophic-Global-Demise-of-Bird-Populations-and-Species.pdf
4. National Geographic. Endangered Species, 2022. Retrieved November 2, 2022 from https://education.nationalgeographic.org/resource/endangered-species
5. Sandraviz. Threatened Species - Birds 2020 Animated, 2021. Retrieved October 2, 2022 from https://observablehq.com/@sandraviz/animated-bar-chart.
6. The BirdLife International. European Red List of Birds, 2021. Retrieved October 2, 2022 fromhttps://www.birdlife.org/wp-content/uploads/2022/05/BirdLife-European-Red-List-of-Birds-2021.pdf.pdf
7. Tyhe Reading. Threatened Species Across the Globe, 2019. Retrieved October 2, 2022 from https://www.behance.net/gallery/85374469/Threatened-Species-Across-the-GlobeData-Visualisation.