

Process Book

Descriptions, concerns and changes:

For the overall visualization project, the challenges arose from the very beginning when we were given the task of finding a database to present and visualize. There were several choices we put forward at the very beginning but finally were rejected due to some reasons. To be specific, we had basically seven ideas:

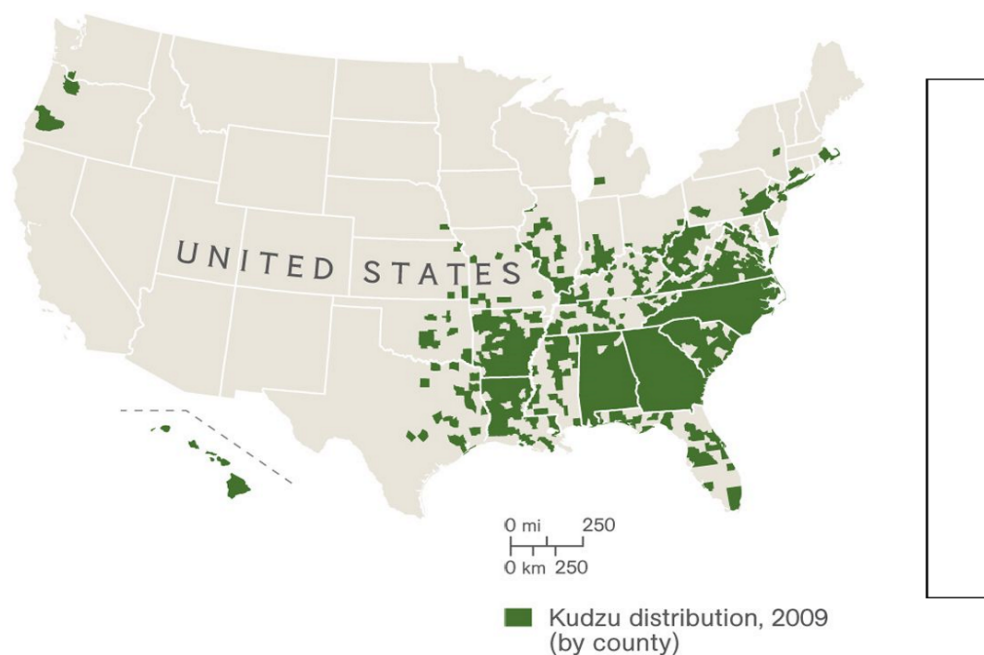
1. Japanese Manga in Europe or other regions other than Japan;
2. visualization about covid data;
3. the history of Switzerland, or Lausanne to be more concentrated;
4. extreme climate change;
5. bird distribution and migration;
6. publication situations before and after French Revolution;
7. the shape of eggs, chairs or mines and so on.

The 1st and 6th ideas were very interesting but they lacked the necessary database. For the 6th idea, even if Professor Roger Chartier provided information about the data he used in his book, we found them quite scattered and hard to visualize completely. The 2nd and 4th ideas had the fullest database with complete pre-process but it was hard to think of novel visualization format and stories about them. The 3rd idea was a quite balanced situation since we've found the huge database on the Swiss government's online pages about economics, populations, etc. for various visualizations. But the problem was that we needed to work heavily on the data process and analysis schema before we could do the data visualization procedure, which is not the purpose of this course. For the 7th idea, it was a little bit difficult to search the shape database of chairs. For mines and eggs, there were already complete data visualizations so it was hard to put forward novel stories. We finally decided to visualize the 5th idea since we had the whole dataset of bird distributions and their corresponding time period with different species. On the other hand, although there were some bird migration visualizations, we were still able to improve them with interactive figures and a more complete database.

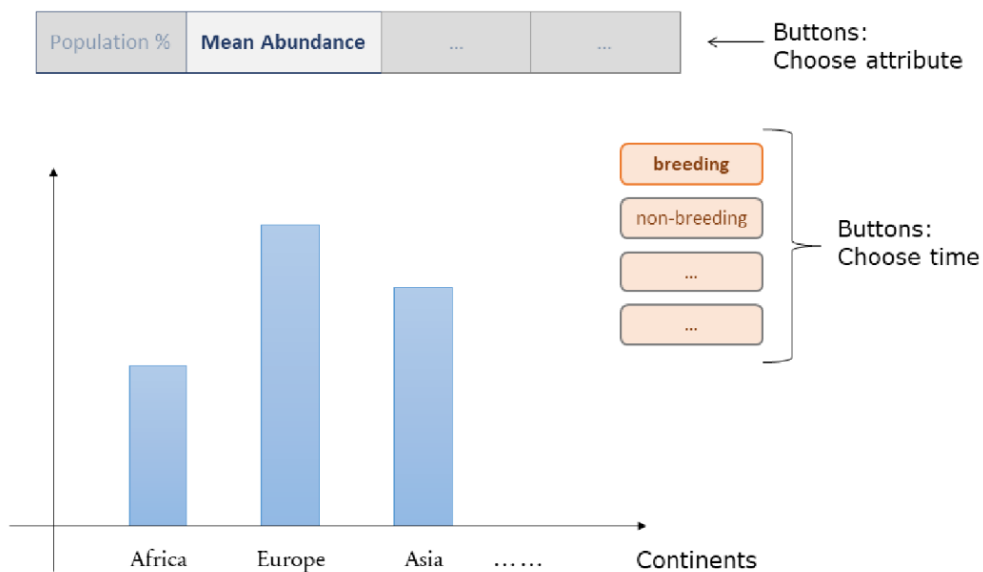
After we decide overall what we will present on our website, the question of how and what exactly we should present arose. By examining the

database of birds on hand, we found that the dataset mainly described the abundance, population and area occupation features of different species within different regions (countries or cities) within different periods. Therefore, our pipeline should mainly present these statistical features together with geometric distributions of birds, which led to two main sections about our project.

In the first section, we want to present a general worldwide geometric distribution of a specific kind of bird. This visualization should be interactive in a specific region with a color bar indicating the density of the bird and easy for the user to track the distributions. We also added a time selection bar so that one could drag the button to see the variation of distributions against different periods. The transition is made to be continuous so that it has an effect on bird migrations. The figure here shows the sketch of our map.

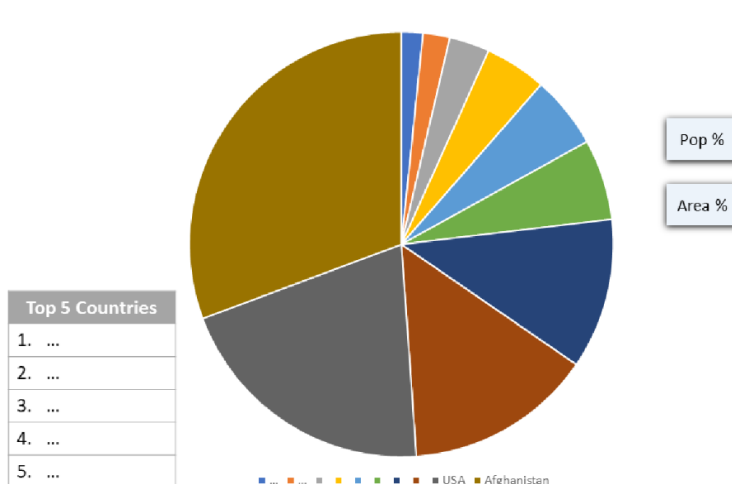
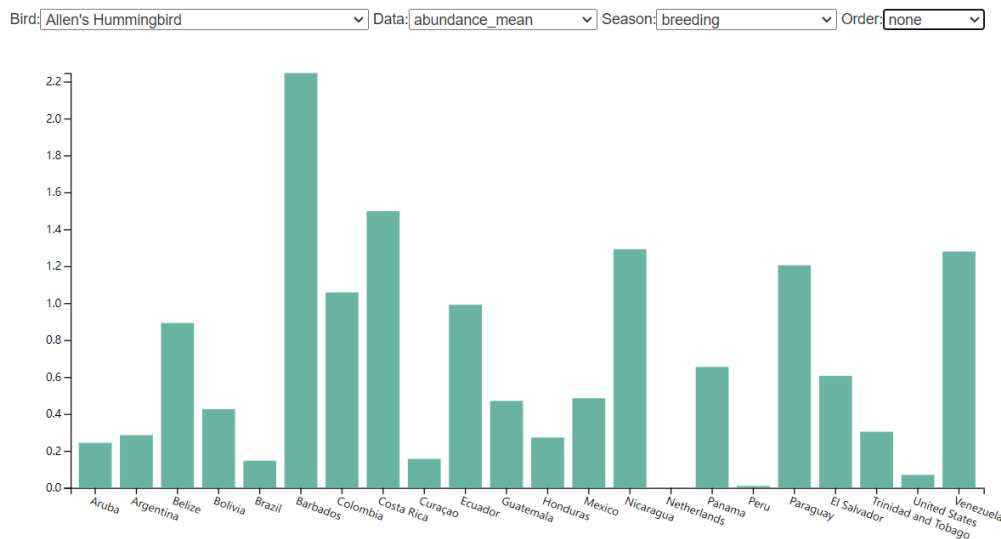


For the second section, we plan to present more detailed statistics of bird distributions by bar and pie chart. The format remains a problem but we finally decided to present features one by one rather than present them simultaneously in a bar chart since the latter format has no advantage to show features of bird distributions and sometimes can even be misleading.



As shown above, we decided to make a selection box for users to select the features they want to learn. In order to include different time periods, we want to add another time selection box so that users can choose between different periods. In the implementation phase, we add more species into our figure to be more abundant. We also modify the selection box so that it's more user friendly now.

However, we noticed that our original proposal was problematic since for one bird, their distributions tend to gather in one or two continents while they are mostly absent in other continents, which leads to the large blank of our figure and less informative. To solve this problem, we add a click event to the bar so that users can access the distributions against countries within a continent by clicking the bar. Further, one can access the distributions within cities of a country by clicking the bar as well. If one would like to return to the previous bar chart, he can click the blank part. Hence, this format of visualization becomes more informative and can provide more information about bird statistics depending on the needs of users.



Although the bar chart is very common and also useful for informative visualization, this kind of format is still a little bit boring and cannot show the percentage and rank of distributions between different

regions. In our previous plan, we intended to supplement the bar chart with a pie chart so that the percentage and rank relation is clearly presented. The motivation behind this is that maybe some users would like to learn the difference of distributions between different regions in percentage forms. For example, one would like to know what's the top five regions that have the most bird numbers of a bird and what's the gap between the first and second as shown below. Again, the pie chart could also have click events so that users could focus on countries, cities or continents. Time selection bar is also added for different time periods.

But in our final implementation, we dropped this idea because it might be too redundant for our visualization. Rather, we noticed that the above visualization procedure weighted too much on one bird's statistics. Hence we put our efforts more on other aspects to visualize our dataset.

[illegible]

Finally to complete our visualization schema, we plan to add one more section where

All Birds: American Golden-Plover Selected Birds: American Golden-Plover reset

Common Name

country

season_name

abundance_mean

total_pop_percent

range_percent_occupied

range_total_percent

American Golden-Plover

American Redstart

American White Pelican

Mexico

Panama

United States

breeding

nonbreeding

postbreeding_migration

prebreeding_migration

0.65

0.60

0.55

0.50

0.45

0.40

0.35

0.30

0.25

0.20

0.15

0.10

0.05

0.00

More detailed difficulties and challenges:

Geographic map:

1. At the beginning we would like to display the bird distribution for every city, but we did not find a dataset containing all the geographic data of cities mapped to their countries. So we decided to show the distribution by countries on the map.
2. Since we noticed that the seasonal distribution of many bird species is incomplete, we decided to consider and to display only the species which have complete data.
3. Our original dataset file was very large, and it took too much to load. So we filter the dataset and keep only the necessary features.

Statistic features:

1. Initially we planned to make a histogram and a pie chart separately. The histogram was expected to plot the distribution by continents, and the pie chart should show the distribution among countries and cities. Then we see that the presence of many birds are limited to a few continents, sometimes they can be found just in a specific continent. So we decided to use only the histogram, with the rectangles representing countries. Clicking on them, the distribution among the corresponding cities is shown.
2. We notice that some bird species are present in many countries, in this way the concentration of the labels on the horizontal axis is too dense. So we add the zoom feature to our graph for a better readability of the labels. The labels are also rotated to avoid overlapping.

Region Distribution & Birds Comparison:

1. We planned to make a sunburst graph to show the distribution at each geographic level, but our dataset was too large and it cannot be rendered. So we choose to make a circular graph with drop-down selection menus for feature selection.
2. We saw that displaying in a circular way affects the readability of a great part of species' names on the graph. So we add a rotating

feature to the graph, allowing the users to rotate it and to read all the names.

Technical implementation:

1. We also wanted to realize the comparison of the bird species distribution among cities. Unfortunately the birds can be found in a huge number of cities and the graph was a mess. We tried to solve the problem by adding the zoom effect to the axis, but we didn't succeed in realizing it. In the end we decide to show the comparison among countries instead of plotting by cities

Supplement database:

1. For the general purpose of our visualization, we showed the bird distributions and its changes within one year. Further, we would also like to collect more data on different years in order to observe the variation of different years. However, even though we have contacted the Cornell bird lab requesting additional data for different years, we are unable to get access to them. Hence, the data in 2020 is all we have and the website is the best we could show.

Peer assessment:

Qiming Sun found the suitable datasets and proposed the idea of plotting a geographic map. He worked on the implementation of the map and on the CSS styling part of the website. He also contributed to the realization of the process book and the screencast.

Hanyuan Hu participated in the selection of the project topic and the datasets. In the early stage, he was committed to the data processing. Then he suggested to implement the statistic features and worked on them. He proposed the styling and the animation effects of the website, and he contributed to the realization of the process book and the screencast, too.

Zewei Xu took part in the project topic selection and in the data processing. He mainly worked on the technical implementation, in particular about the region distribution and the bird comparison. He participated in the realization

of the screencast and proposed the issues for the process book's challenges summary.