VAST - Mini challenge 1

Team Nisaguru

November 2018

Contents

Introduction

The VAST Mini challenge gives a dataset of observations on a specific region. The data is big and you need a tool to visualize it. There are two properties of the dataset that make difficult its observation. First, the data is about a lot of birds with nineteen species registered. Second, the data is sparse and through a long period, 35 years. This paper describes how we solved this issues and what we have found.

1 Showing the observations on a map

The datasets contains two thousands observations of birds from 1983 to 2018. In 35 years, 19 species have been spotted. The observations have been made in a specific region and VAST only gives an image of it.

With 35 years of data, it is not possible to plot every observations in just one figure. Indeed, you only get a lot of point everywhere and it does not make a lot of sense. At least, you can see where the birds have been through the 35 years but no patterns. To show the movement of population we choose to build an animation. Like a replay of what happened in the region.

To do so, the first thing was to choose the step of our animation. As the observations are stored at a day-precision, we choose to use a step of one day. However, with just one day, the observation are not shown enough time to be seen by the user. It only makes the observation twinkle very fast and it is not useful. The solution was to use a window of observation. The observations are not shown just one day but from $T \pm L/2$ days. Then, they are shown for L days and now it is visible by the user. We choose to use a centered window because we suppose that the bird was there before and after the observation. At least for its life span. To simulate the apparition and the extinction of the bird, we used a fade of the alpha property. The user can now see the birds show and disappear from the screen. It makes the animation a lot smoother.

Another issue with the dataset was that there are a lot of species. Then it is difficult to see which one is what. To solve it, we choose to use a categorical color palette. This palette tries to find the best color to show for 19 species in one plot. However this is not enough, nineteen is too much to be understood clearly. Then we choose to give the user the choice of which species would be shown in the animation. With a simple select box, the user can choose at anytime the species he wants to see. This enable the user to only see 2 or 3 species that could behave together while not being polluted by the 16 others species.

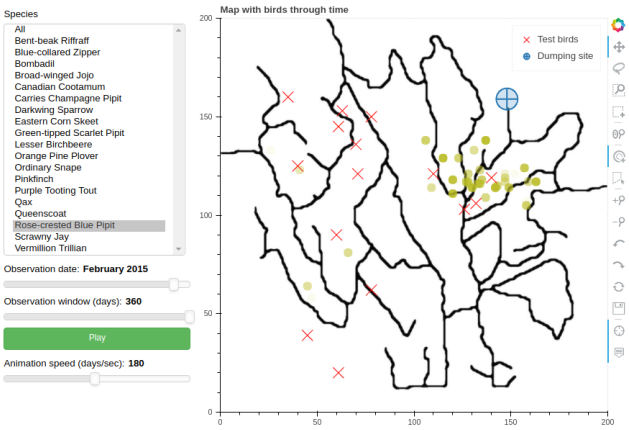

Figure 1: The user interface of the animation.

You can see the animation on figure 1. This animation is useful because it enables the user to see any species at any time in the region. With such tools, you can spot nearly every pattern very quickly. For example, you can clearly notice that the Blue Pipit are migrating from the location of the dumping site of Kasios waste products.

So the animation gives a lot of liberty to the user but it has one drawback : the user does not know where to start. This animation has been made to visualize patterns in the data. When you have found one, you should analyze it with more specific tools.

2 Pattern analysis

To analyze the migration of Blue Pipit, we have used the kernel density over the years.

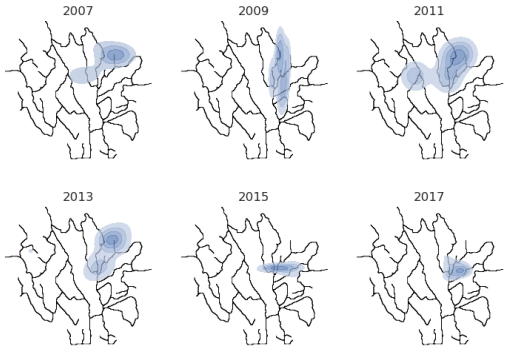

Figure 2: The kernel density over the years of Blue pipit.

You can see on figure 2 that the Blue Pipit are moving over the years. It seems that since the dumping site was built, they moved to the South. However, it is difficult to say if its population has decreased. Indeed, there are very few observations in the beginning of the period and a lot more in the end. We cannot be sure that the observation are done uniformly. Then it is difficult to say if the increasing number of observation is due to more Blue Pipit or just to more people looking for it.

The kernel density is useful for a report. It gives a static view of what happened. The animation above cannot be put nicely in a report. It is a tool to explore the data.

Conclusion

The animation is useful to spot quickly any pattern you want about any bird. However, it fails to give a precise view of a specific pattern. With a specific pattern in mind, it is better to use the kernel density than the animation.

This two tools enabled us to find that the Blue Pipit were moving from the dumping site.