

Avaliação de Padrões de Migração de Aves Utilizando Dados de Repositórios Públicos de Ciência Cidadã

Juliana Rezende S. B. Alves¹, Jéssica Naiara Reis², Pedro Arthur R. L. Silva¹
Leonardo Esteves Lopes², José Augusto M. Nacif¹

¹Instituto de Ciências Exatas e Tecnológicas, Campus UFV-Florestal
Rodovia LMG 818, km 06 – Florestal – MG – CEP – 35690-000 – Brazil

²Laboratório de Biologia Animal, IBF, Campus UFV- Florestal
Universidade Federal de Viçosa, Brazil
Rodovia LMG 818, km 06 – Florestal – MG – CEP – 35690-000 – Brazil

{juliana.baia, jessica.naiara, pedro.arthur, leonardolopes, jnacif}@ufv.br

Abstract. *The lack of data has always been an obstacle for the understanding of the birds migratory movements, especially in the tropics. The recent development and increasing popularity of various projects of Citizen Science have provided a great set of data to the scientific community, which has grown exponentially. Such data have been collected and tabulated by people without academic training, thus require a distinguished management and validation, because, otherwise, they can lead to inaccurate conclusions. The study of migratory patterns of certain birds species is a good example of the use of Citizen Science, because birdwatching's increasing popularity in tropical regions has allowed a data accumulation of species and regions that, so far, have never been satisfactorily sampled by the scientific community. The purpose of this paper is to use the Data Science and Citizen Science conjointly, in order to evaluate the reliability of the data that have been made available by two datasets with distinct characteristics, one based in vouchers (WikiAves) and the other based in undocumented observations (eBird). We have studied the migratory movements of four different birds species, with different migration patterns: Amazona pretrei, Attila phoenicurus, Casiornis fuscus and Tangara peruviana. Our results show that the WikiAves dataset, after analysis, was more satisfactory and consistent with the migratory patterns, compared to the eBird.*

Resumo. *A carência de dados sempre foi um obstáculo para o entendimento dos movimentos migratórios das aves, especialmente nos trópicos. O desenvolvimento recente e a crescente popularidade de diversos projetos de Ciência Cidadã disponibilizaram para a comunidade científica um enorme conjunto de dados, que tem crescido exponencialmente. Tais dados, por terem sido coletados e tabulados por pessoas sem treinamento acadêmico, requerem gerenciamento e validação diferenciados, pois, caso contrário, podem levar a conclusões equivocadas. Um bom exemplo de uso da Ciência Cidadã tem sido o estudo dos padrões migratórios de espécies de aves, pois a crescente popularidade do birdwatching em regiões tropicais tem permitido um acúmulo de*

dados sobre espécies e regiões até então nunca amostradas satisfatoriamente pela comunidade científica. O objetivo deste trabalho é utilizar a Ciência dos Dados conjuntamente com a Ciência Cidadã para avaliar a confiabilidade dos dados disponibilizados por duas bases de dados com características distintas, uma baseada em Vouchers (WikiAves) e outra em observações não documentadas (eBird). Para isso, foram investigados os movimentos migratórios de quatro espécies de aves com distintos padrões de migração: Amazona pretrei, Attila phoenicurus, Casiornis fuscus e Tangara peruviana. Concluiu-se que a base de dados WikiAves, após análises, se mostrou mais satisfatória e condizente com os padrões, comparada ao eBird.

1. Introdução

O estudo da ecologia da movimentação em aves é uma tarefa árdua, pois elas exibem uma infinidade de diferentes tipos de movimentos, muitos deles extremamente complexos. Historicamente, estudos de migração têm sido feitos com base na captura, marcação (anilhamento) e recaptura, associadas à avaliação de variações nos padrões sazonais de abundância ao longo da distribuição geográfica de cada espécie. O problema desses dois principais tipos de abordagens é que, primeiro, a recaptura de uma ave marcada é um evento raro e, segundo, a obtenção de dados de ocorrência é extremamente trabalhosa, exigindo a visita a dezenas de coleções zoológicas do Brasil e exterior, detalhada revisão de literatura e a revisão de dados não publicados coletados por outros cientistas. O resultado, mesmo após este esforço, é um grande enviesamento dos dados em direção a algumas espécies e regiões mais bem estudadas, com grandes lacunas de dados para espécies e regiões difíceis de serem amostradas.

Felizmente, nos dias de hoje, existe uma pluralidade de maneiras de se adquirir conhecimento e coletar informações. A fim de otimizar o tempo da coleta de dados, tem se tornado cada vez mais frequente a utilização de dados coletados por leigos voluntários. A colaboração entre cientistas profissionais e leigos é conhecida como Ciência Cidadã [Resnik et al. 2015]. Voluntários ajudam na coleta de dados e interagem com cientistas a fim de ajudar em alguma causa científica ou, muitas vezes, colaboram com os registros como uma forma de hobby. Dessa forma, o que começa como uma distração, pode vir a se mostrar de valor científico. Contudo, como tais voluntários podem ser pessoas leigas no assunto, a coleta e/ou tabulação dos dados pode ocorrer de forma equivocada, prejudicando a qualidade e a confiabilidade dos dados disponibilizados. Por conseguinte, esses equívocos se tornam ruídos, levando em conta que muitas vezes os sites não fazem um pré-processamento dos seus dados ou não possuem políticas de confiabilidade.

Em virtude da quantidade de dados gerados pela Ciência Cidadã, é necessário um método para lidar com um volume tão grande de dados. Por esse motivo, a Ciência dos Dados tem se mostrado uma grande aliada da Ciência Cidadã, pois permite a aplicação de métodos para resolver problemas relevantes e prever resultados [Waller and Fawcett 2013]. Com a Ciência dos Dados, é possível fazer uma limpeza dos ruídos nos registros, agindo como um tipo de coletor de dados importantes os quais são fundamentais para facilitar, neste caso, na obtenção dos registros sobre a migração das aves aumentando a confiabilidade dos dados. Além disso, com a Ciência dos Dados também é possível fazer diversas análises que nos possibilitam fazer uma comparação

entre as bases de dados, permitindo, por exemplo, examinar qual das bases de dados estudadas é a mais confiável.

Para o estudo das aves tropicais, duas bases de dados de Ciência Cidadã merecem destaque. São estas o eBird [eBird 2012] e o WikiAves [WikiAves.com 2008]. O eBird foi especificamente projetado para ser utilizado como uma plataforma de Ciência Cidadã em nível mundial, permitindo ao usuário o download simples e rápido do seu banco de dados em um formato analisável. O Wikiaves, por outro lado, não foi projetado originalmente para ser uma plataforma de Ciência Cidadã, mas sim um site de conteúdo interativo, voltado para a comunidade brasileira de observadores de aves. Dessa forma, o WikiAves não permite o download de seus dados sendo necessário a implementação neste trabalho de um Scraper, método para extrair dados de páginas web, para a obtenção dos mesmos. Uma diferença marcante entre os dois é que o WikiAves utiliza de Vouchers digitais, que são fotos e gravações de vocalizações as quais permitem a validação dos dados de maneira independente por terceiros. Os Vouchers também são considerados a partir do momento em que o usuário interage com os registros dando curtidas, positivas ou negativas. Já o eBird é amplamente baseado em registros observacionais desacompanhados de Vouchers, o que inviabiliza a validação posterior do registro, o que pode ser um problema para espécies de identificação complexa.

Neste trabalho, foi feita uma comparação entre a confiabilidade dos dados disponibilizados por estes dois sites e a sua eficiência para identificar os padrões migratórios de quatro espécies de aves cujos movimentos migratórios já se encontram bem conhecidos. A investigação foi feita através da análise de gráficos de dispersão gerados através dos registros dos sites que nos revelaram aspectos importantes sobre os movimentos migratórios das aves.

2. Background

2.1. Ciência Cidadã

A História Natural sempre foi associada com a ciência [Turnhout et al. 2016] e os chamados voluntários sempre tiveram grande participação no seu desenvolvimento. A participação dos voluntários se dava principalmente por meio de registros que eram submetidos e, assim, era estabelecida uma comunicação de forma comunitária. A motivação dos voluntários era, não somente ajudar a ciência, mas o estudo da História Natural em si. Sendo assim, este tipo de coleta de dados é intitulada de Ciência Comunitária, na qual os voluntários interagindo de forma colaborativa produzem o conhecimento e tendem a produzir algum tipo de mudança.

Já a Ciência Cidadã se difere da Ciência Comunitária pois nela o avanço da ciência está em primeiro lugar e não são os voluntários quem dirigem os estudos. Na chamada Ciência Cidadã, os voluntários são chamados de cidadãos e há uma relação forte entre eles e os cientistas. As primeiras definições que devemos entender são: o que é o cientista e o que é o cidadão, neste contexto, para um melhor entendimento. O cidadão se difere do cientista pelo fato de não ser um pesquisador profissional e com treinamento formal [Resnik et al. 2015], mesmo podendo ter algum tipo de experiência no assunto pesquisado ou tratado no projeto em questão. São voluntários que participam como assistentes em pesquisas e estudos científicos [Cohn 2008]. Geralmente são pessoas que têm interesse pelos assuntos pesquisados, preocupadas com o propósito biológico envolvido e que

gostam de estar no campo coletando estes dados [Cohn 2008]. Já o cientista é o profissional formado em certa área específica, envolvido em projetos na área de seu interesse e que está em contato com o cidadão nesta troca de informações.

A Ciência Cidadã é considerada uma nova forma de fazer ciência [Cohn 2008], pois os dados são coletados de uma maneira completamente diferente e inovadora, a qual resulta na obtenção de um grande volume de dados em um curto espaço de tempo. É praticamente impossível coletar a mesma quantidade de dados, com uma cobertura geográfica tão ampla, utilizando as técnicas tradicionais de pesquisa. Isso porque é economicamente, tecnicamente e logisticamente impossível enviar um número equivalente de técnicos e especialistas ao campo por um período de tempo tão curto.

A Ciência Cidadã traz a ideia de que, juntos, cientistas e cidadãos podem mais [Ramesh et al. 2017], uma promessa que contribui decisivamente para que esta técnica de coleta de dados seja tão eficiente. O fato do cidadão sentir que a sua colaboração ajuda na conservação da biodiversidade ou em alguma outra causa na qual acredita faz com que o mesmo continue sempre registrando e colaborando com os projetos. Caso o cidadão sinta o contrário, que não está colaborando, ele pode vir a parar de colaborar [Turnhout et al. 2016].

Um ponto, porém, recorrente quando se fala em Ciência Cidadã é em relação à confiabilidade dos dados. Isso porque os dados são coletados primariamente por pessoas leigas, sem treinamento científico, que podem acabar por coletar dados erroneamente ou mesmo entrar com os dados na plataforma de maneira equivocada. Tais erros podem corromper o conjunto de dados tornando-o pouco confiável. A validação dos dados tem que ser feita no momento em que a observação é transformada em registro. A partir do momento em que muitos dados são enviados é necessária uma validação [Turnhout et al. 2016]. Esta validação antes não era tão necessária pois os portadores do conhecimento eram pessoas treinadas ou até mesmo técnicos no assunto. Mas agora, com a introdução de possíveis ruídos nas bases de dados é de extrema importância a vistoria dos registros a fim de se manter a confiabilidade dos mesmos.

Apesar dessa desvantagem, este tipo de coleta de dados está se mostrando forte e no Brasil já é possível encontrar projetos baseados em Ciência Cidadã e no seu conceito. Esses projetos possuem enfoque em áreas como a biodiversidade e a conservação de espécies. Fora do Brasil, o uso da Ciência Cidadã também vem se ampliando. Após o exposto, serão descritos dois projetos baseados em Ciência Cidadã nas Subseções a seguir.

2.2. WikiAves

Criado em 2008, o Wikiaves é uma iniciativa isolada de um grupo de observadores de aves da Sociedade Civil que tem como objetivo principal apoiar e divulgar a atividade de observação de aves [SiBB(1)]. Entretanto, o volume e a qualidade de dados disponibilizados pelo WikiAves faz com que o site venha sendo utilizado de maneira crescente pela comunidade científica, ainda que o seu propósito original não tenha sido esse. Prova disso é que o Wikiaves é um dos sites de maior destaque dentro do Sistema de Informação sobre a Biodiversidade Brasileira (SiBB) [SiBB], uma iniciativa do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) que objetiva a obtenção de dados e conteúdos em biodiversidade através de plataformas online. Atualmente o projeto recebe

mais de 300.000 registros de aves brasileiras por ano [SiBBR(1)].

O WikiAves utiliza um sistema de Voucher Digital que aumenta moderadamente a confiabilidade dos dados. Vouchers digitais podem ser considerados desde o próprio registro do cidadão, que no caso do WikiAves seriam as fotos e gravações de vocalizações das aves, até mesmo a interação do mesmo com o site, que é o caso do usuário do referido projeto. Por ser um site bastante acessado, oferece interface interativa e permite que o usuário curta fotos e registros de aves. Dessa forma, se por um acaso um registro errado de alguma ave passar no sistema, existe a possibilidade de outro usuário estar alerta e classificar aquele registro como errado ou duvidoso, aumentando assim a confiabilidade do site, pois ao acessá-lo, outro usuário, ao notar uma curtida negativa, irá perceber que aquele registro provavelmente está classificado erroneamente. A revolução dos "Vouchers Digitais" estabeleceu uma era na Ornitologia qualitativa, baseada em evidências, que permite uma revisão instantânea de registros de aves e uma expansão do número total de registros como observações ornitológicas amadoras. [Lees and Martin 2015]. Os Vouchers, aliados à uma boa validação de dados, com certeza trazem um amplo grau de confiança para qualquer conjunto de registros.

2.3. Ebird

O eBird [eBird 2012] surgiu com a proposta de documentar os registros dos observadores de pássaros com o intuito de compartilhar o conhecimento posteriormente. O site fornece um banco de dados de ocorrência de aves cobrindo amplas escalas espaciais. [Ramesh et al. 2017]. É um dos maiores projetos de biodiversidade do mundo baseado em Ciência Cidadã [eBird 2012], envolvendo milhões de dados que são disponibilizados para estudos e análises.

A proposta é parecida com a do WikiAves, exceto pelo fato de que o eBird não documenta somente pássaros brasileiros e que não há Vouchers Digitais dos Cidadãos, pois o site não é totalmente interativo, apresentando cunho mais científico. Apesar de não possuir Voucher Digital, o eBird é um projeto científico muito utilizado em pesquisas, especialmente para espécies Norte Americanas, sem maiores problemas. Um dos motivos seria a cultura de observação de aves nos Estados Unidos que é muito difundida, com milhões de observadores amadores contribuindo com dados. Além do número de pessoas, outras diferenças importantes existem entre a comunidade de observadores de aves dos EUA e do Brasil: 1) os observadores norte-americanos geralmente possuem equipamentos (binóculos, lunetas) e guias de campo de melhor qualidade; 2) a biodiversidade dos Estados Unidos é muito menor do que a brasileira; 3) a tradição de observar aves é muito antiga nos EUA, atravessando gerações de uma mesma família, enquanto que no Brasil é uma atividade extremamente jovem. Todas essas diferenças fazem com que o eBird apresente problemas com a validação de seus dados no Brasil, sendo o seu uso como ferramenta científica questionado para espécies tropicais [Lees and Martin 2015].

3. Metodologia

Nesta seção, serão apresentadas as espécies estudadas e uma descrição detalhada do padrão migratório de cada ave descrito pela literatura. Logo após, será discutido como foram obtidos os dados utilizados para a realização deste trabalho e, em seguida, será abordado como foi feita a análise dos dados para a obtenção dos resultados.

3.1. Espécies estudadas

As espécies de aves analisadas neste trabalho foram: *Amazona pretrei*, *Attila phoenicurus*, *Casiornis fuscus* e *Tangara peruviana*. Tais aves foram escolhidas por diferentes motivos: 1) elas exibem movimentos migratórios já razoavelmente bem conhecidos, existindo, portanto, hipóteses bem estabelecidas na literatura sobre o padrão migratório de cada uma delas [Somenzari et al. 2018]; 2) elas exibem diferentes padrões migratórios, proporcionando, portanto, oportunidades distintas para se testar de maneira independente a eficiência dos dados de ciência cidadã para revelar o correto padrão de movimentação de cada espécie; 3) elas possuem diferentes graus de complexidade de identificação, o que permite testar a confiabilidade dos dados de ciência cidadã para espécies de identificação complexa; e 4) as quatro espécies aqui estudadas apresentam a quase totalidade da sua área de distribuição restrita ao território nacional, o que é de fundamental importância para este estudo, já que o WikiAves cobre somente o território brasileiro, diferentemente do eBird, que é uma base de dados de cobertura global. Portanto, a escolha de espécies com distribuição geográfica altamente coincidente com as fronteiras brasileiras é imperativa para que seja lícita a comparação entre as duas bases de dados aqui avaliadas.

- *Amazona pretrei*: migrante de curta distância. A espécie se reproduz de meados de julho a meados de janeiro em uma ampla área de reprodução nos estados de Santa Catarina e Rio Grande do Sul. Depois eles migram para locais de concentração populacional, onde se congregam em dormitórios coletivos pós-reprodução, permanecendo nessas áreas até março [Martinez and Prestes 2002]. De abril a meados de julho eles se dirigem para os sítios de invernada, onde se alimentam dos pinhões das araucárias (*Araucária angustifolia*), nos últimos remanescentes de floresta de araucária na divisa de Santa Catarina e Rio Grande do Sul [Somenzari et al. 2018].
- *Attila phoenicurus*: Entre 15 de outubro a 15 de março é o período de reprodução da espécie, principalmente na Serra do Mar do litoral sudeste do Brasil, nos domínios da Mata Atlântica. De 15 de março a 15 de abril e de 01 setembro a 31 outubro, observa-se o período transiente, quando a espécie se desloca pelo Brasil Central e nordeste extremo da Argentina, leste do Paraguai e Bolívia, até atingir o sítio de invernada. De 15 de abril a 1 de setembro, *Attila phoenicurus* encontra-se invernando em uma ampla área da Amazônia Brasileira, atingindo também o sul da Venezuela [Somenzari et al. 2018]. Até onde se sabe, todos os indivíduos da espécie realizam migração. Esta espécie é a que apresenta maior dificuldade de identificação em campo, pois habita o dossel de florestas úmidas, onde as condições de iluminação são precárias, sendo, por isso, facilmente confundida com fêmeas de *Pachyramphus validus* por observadores inexperientes.
- *Casiornis fuscus*: Esta espécie é migrante parcial, pois nem todos os indivíduos de uma população realizam movimentos migratórios. Os migrantes exibem o seguinte padrão de deslocamento: de dezembro a março *Casiornis fuscus* se encontra em seu período de reprodução, que se concentra na Caatinga do nordeste brasileiro e em suas áreas de contato com as províncias biogeográficas vizinhas. De junho a setembro se encontra invernando no leste da Amazônia e nas complexas áreas de transição com o Cerrado. Em abril-maio e outubro-novembro a ave se encontra no período transiente, deslocando-se entre os sítios de reprodução e invernada [Somenzari et al. 2018].

- *Tangara peruviana*: Migrante parcial, habitante das restingas e florestas úmidas do litoral sudeste e sul do Brasil. De 15 de setembro a 15 de janeiro, se encontra no período de reprodução, que se dá na porção centro-sul de sua área de distribuição. De abril a setembro parte da população de *Tangara peruviana* se desloca para o extremo norte de sua área de distribuição, nos estados do Espírito Santo e Rio de Janeiro, onde inverna [Somenzari et al. 2018].

3.2. Obtenção dos dados

Para a obtenção dos registros do site WikiAves [WikiAves.com 2008] foi implementado um Scraper pois o site não fornece os dados para estudos/pesquisas e nem opções para exportação dos dados. A técnica de Web Scraping é muito utilizada para obter dados de páginas da web e extrair informações úteis para a análise de dados. Tal método pode ser implementado de diversas formas e utilizando diversas bibliotecas. O Scraper em questão foi implementado em Python, utilizando a biblioteca BeautifulSoup¹ como auxiliar e outras bibliotecas.

O primeiro passo do algoritmo consiste em formar os links que correspondem a cada registro de aves no site. O link para acessar essas páginas segue um padrão, que é formado de um link base e o id do registro acessado. O link base é o wikiaves.com.br e o id é um número inteiro o qual começa a partir do 0. Dessa maneira, o link completo para se ter acesso a um registro desse site tem a seguinte forma: www.wikiaves.com.br/id.

A próxima etapa do Scraper implementado consiste em utilizar os links formados para recuperar as páginas HTML que contêm as informações necessárias para nossa análise. Para isso, utilizou-se a biblioteca requests² do Python, que serve para realizarmos requisições HTTP utilizando apenas comandos do próprio Python. Com base somente nos links que formamos na etapa anterior, podemos utilizar a requests para realizar uma requisição que recuperará o HTML da página correspondente ao endereço de página atual.

O passo seguinte é analisar o HTML recuperado e extrair as informações desejadas da página. Para isso, foram utilizadas expressões regulares as quais buscam na página HTML do site as informações necessárias para nosso estudo, que no caso foram os campos: Nome da espécie, Local da observação (Município), Estado, Data na qual a foto foi feita, Autor e link de acesso. Foi possível implementar desta forma pois as páginas de registro de cada ave do site seguem um padrão. Dessa forma, sempre vão haver estes campos nas páginas, as alterações nos campos já são previstas e implementações para tais alterações já implementadas. Após localizar os campos, é criado um arquivo .CSV o qual é usado para manipulação na análise de dados. O fluxograma visto na Figura 1 nos mostra de forma mais explicativa como foi implementado.

O Scraper irá realizar a busca de páginas, que representam as aves, até o momento em que sejam encontradas mais de 500 páginas seguidas que não contenham nenhum registro. O critério de parada foi feito dessa forma, pois dentre os registros normais do site existem alguns que foram excluídos pelos próprios usuários, sendo assim é normal que o algoritmo encontre algumas páginas seguidas sem nenhuma informação, mas altamente improvável que encontre um número alto de páginas sem registros continuamente.

¹<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

²<https://pypi.org/project/requests/>

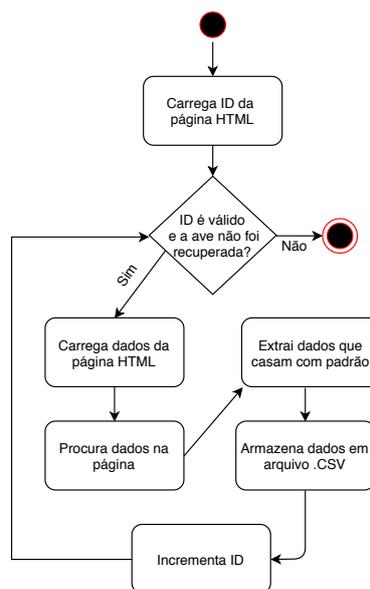


Figura 1. Fluxograma referente ao Scraper implementado, utilizado para a coleta de dados do site WikiAves.

Já para a obtenção dos registros do site eBird [eBird 2012] não foi preciso fazer nenhuma implementação à parte para a extração dos dados. O site já fornece os arquivos com todos os dados necessários para a análise de dados e eliminação de ruídos, necessitando apenas criar uma conta no site e pedir acesso aos registros. O que faz todo sentido, posto que o eBird é um site de cunho científico voltado para pesquisas.

Além disso, para este trabalho eram necessárias as coordenadas (latitudes e longitudes) das cidades dos registros provenientes do WikiAves. Porém, o site não fornece este campo em sua página web, sendo necessário uma conversão do nome das cidades coletadas, através da implementação explanada acima, a qual foi feita utilizando a biblioteca GeoPy³. Tal biblioteca define um método reverso que resolve um par de coordenadas para um endereço, ou seja, dado um nome de cidade ela nos retorna a latitude e longitude respectiva, facilitando no georreferenciamento. Foi utilizada a aplicação web Jupyter Notebook⁴ como IDE pois a mesma possui diversas bibliotecas que ajudam na análise de dados. Tal IDE foi utilizada para a implementação dos gráficos na análise de dados, utilizando a linguagem de programação Python.

3.3. Análise dos dados

Para facilitar a interpretação e a visualização gráfica dos dados, as datas dos registros foram convertidas para JulianDays. O JulianDay recebe como entrada uma data disposta em ano-mês-dia e converte em um número compreendido entre 1 e 366 (ano bissexto), tornando a compreensão e análise bem mais eficiente posto que serão visualizados os dias de um ano completo, podendo observar como uma ave se comporta durante um ano, sem intervalos de tempo cortados. Dessa maneira, 1º de janeiro corresponde ao dia 1, e 31 de dezembro ao dia 365 (366 no caso de anos bissextos).

³<https://pypi.org/project/geopy/>

⁴<https://jupyter.org/>

Foram construídos, para cada espécie, três tipos de gráficos: um gráfico relacionando latitude com os dias, convertidos em JulianDays como já mencionado; outro gráfico para relacionar a longitude com os dias; e por fim, um gráfico relacionando latitude com longitude, levando em conta também os padrões de migração de cada espécie implementados através de uma *colorbar* implementada em Python, na qual cada cor representou um momento distinto do ciclo migratório de cada espécie. A cor fria, representada pela cor azul, indicou a época de invernada; já a cor quente, representada pela cor vermelha, indicou a época de reprodução; e, por fim, a cor alaranjada indicou a época na qual a ave pode estar transitando, indo de um lugar para o outro, ou concentrada em áreas pós reprodutivas (*Amazona pretrei*). Como cada ave segue um padrão de migração diferente foram feitas *colorbars* diferentes para cada espécie.

Primeiramente, avaliou-se a eficiência das duas bases de dados em recuperar os padrões de migração já bem estabelecidos pelas quatro espécies, conforme relatado pela literatura. Em um segundo momento, avaliou-se a confiabilidade das duas bases de dados analisadas, o que foi feito a partir da avaliação da quantidade de ruído inserida nas análises em virtude de registros supostamente incorretos (i.e. registros temporalmente incoerentes com os padrões migratórios das espécies).

4. Resultados e discussão

As Figuras 2, 3, 4 e 5 mostram uma comparação dos dados obtidos para *Attila phoenicurus* utilizando as duas diferentes bases de dados consultadas. Ambas as bases de dados mostram uma marcada concentração de registros da espécie na Mata Atlântica do Sudeste/Sul do Brasil durante o período reprodutivo, o que está de acordo com a literatura. Ambas as bases de dados também mostram escassos registros durante o período transiente e de inverno, o que também está de acordo com o esperado, porque durante essa época a espécie, que geralmente é registrada primariamente por suas vocalizações, permanece silenciosa, não respondendo ao playback do seu canto. A principal diferença entre as duas bases de dados é que o eBird apresenta um número elevado de registros para a Mata Atlântica durante o período de invernada e mesmo transiente, demonstrando que erros de identificação podem estar contradizendo, erroneamente, um padrão migratório bem estabelecido, pois sugerem residência de parte da população. Já na base de dados do WikiAves, poucos são os registros para a Mata Atlântica durante o inverno, alguns dos quais, após checagem individual por um especialista (Leonardo Lopes), revelaram-se erros de identificação com *Pachyramphus validus*, conforme já alertado no início desse estudo.

Quanto aos gráficos de latitude do *Attila phoenicurus*, fica evidente que na Figura 4 temos registros da ave em localizações improváveis dada a época do ano em que elas foram obtidas. Aproximadamente entre os dias 120 e 300 não eram esperados registros da ave na Mata Atlântica, apenas mais ao norte. A Figura 5 é mais condizente com o padrão migratório da ave, levando a crer que a base de dados do WikiAves representa melhor o padrão migratório da espécie.

Já as Figuras 6, 7, 8 e 9 mostram a comparação entre os dados de *Amazona pretrei* obtidos a partir das duas bases de dados consultadas. De acordo com o padrão de migração já bem estabelecido para a espécie, o Wikiaves (Figura 7) mostra claramente que a ave pode ser encontrada no seu período de reprodução, que vai de junho a meados de janeiro

na porção centro-sudoeste da sua área de distribuição. Durante o período pós-reprodutivo a espécie ocupa uma região similar à ocupada durante o período reprodutivo. Já a área de invernada se concentra em uma porção estreita do nordeste da distribuição da espécie. Em contrapartida, a Figura 6 nos mostra que a bases de dados do eBird nos retornou um gráfico com um padrão não tão claro.

Os gráficos de latitude para o *Amazona pretrei*, presentes nas Figuras 8 e 9, nos indica uma similaridade entre as bases de dados e que a ave tem uma área onde ela se mantém residente durante os períodos do dia 50 ao 200, aproximadamente, mantendo uma concentração dos registros ao redor da latitude 28°. Isso se dá pelo fato da espécie passar este período de tempo se alimentando de pinhões próximos a essa latitude, onde se concentram os últimos remanescentes de floresta de araucária, fundamentais para a conservação da espécie. É perceptível que a Figura 9 nos mostra de forma mais coesa e clara a migração latitudinal da ave comparado a Figura 8, nos permitindo analisar o seu comportamento ao longo do ano.

As Figuras 10, 11, 12 e 13 mostram a comparação entre os dados da ave *Casiornis fuscus*. Esta espécie possui ocorrências do leste da Amazônia ao leste do Pernambuco. Como pode ser visto na Figura 11, existem registros da espécie durante o período transitente e de inverno ao longo de toda a área de distribuição da espécie, porém, os registros durante o período reprodutivo estão concentrados fora da região Amazônica, demonstrando que a espécie, que realiza migrações parciais, é um visitante de inverno na região norte do país. Já a Figura 10, não mostra um padrão tão claro.

Como o *Casiornis fuscus* faz migrações longitudinais mais aparentes, o gráfico de latitude relacionado com os dias não nos confere uma análise muito boa. Desta maneira, os gráficos de longitude dispostos nas Figuras 12 e 13 nos reforça o que foi dito anteriormente. Na Figura 13 vemos um padrão claro da ave se deslocando no começo de abril e se dispersando, enquanto na Figura 12, não podemos descrever um padrão claro.

Nas Figuras 14, 15, 16 e 17 podemos ver a comparação entre os dados da ave *Tangara peruviana*. Uma importante diferença observada é em relação à cobertura geográfica dos dados, com a base do Wikiaves mais completa (Figura 15), contemplando toda a porção nordeste da área de distribuição da espécie. Os dados do Wikiaves demonstram que parte da população de *Tangara peruviana* migra durante o inverno austral para o Rio de Janeiro e Espírito Santo, especialmente no período de Abril e Setembro. Já a base de dados do eBird (Figura 14), pela sua menor cobertura geográfica, não demonstra esse padrão de maneira tão clara. O gráfico de latitude desta espécie demonstra bem esse padrão, como podemos ver nas Figuras 16 e 17.

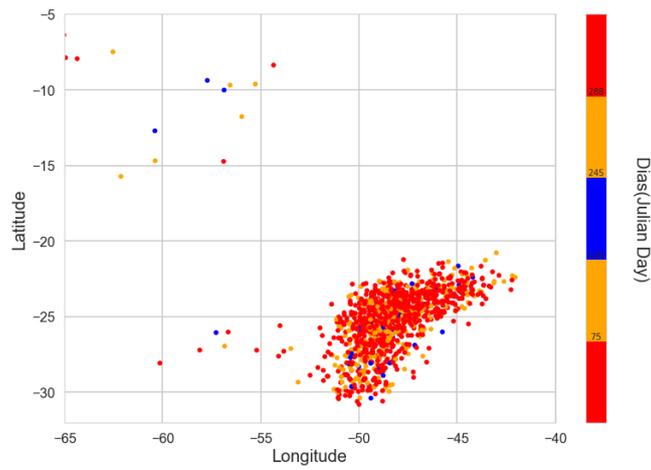


Figura 2. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Atila phoenicurus* retirado do site eBird.

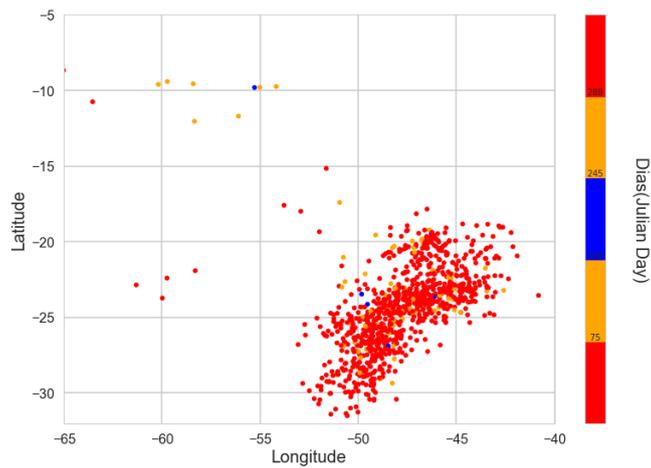


Figura 3. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Atila phoenicurus*, retirado do site WikiAves.

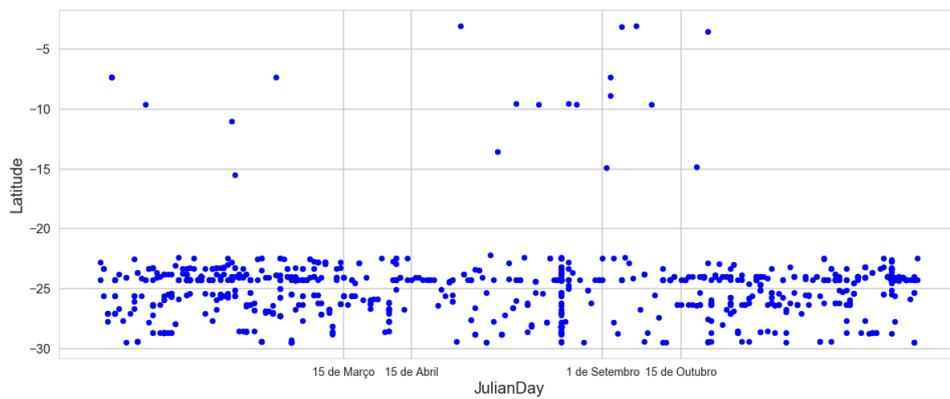


Figura 4. Gráfico de dispersão de latitude da espécie *Atila phoenicurus* retirado do site eBird.

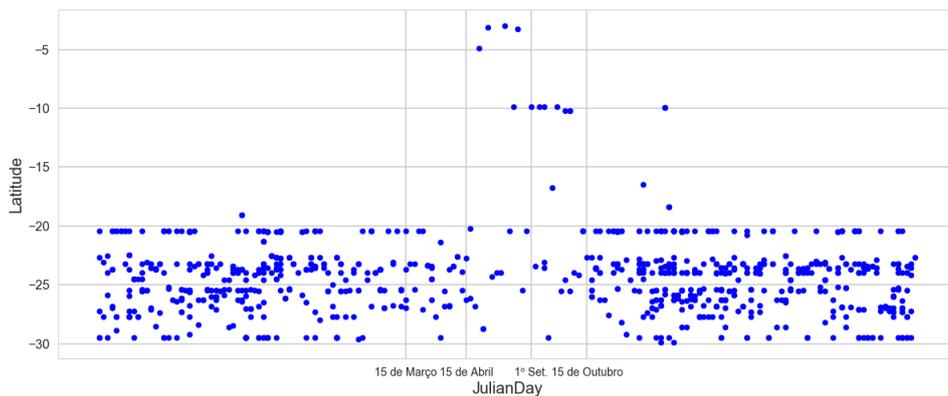


Figura 5. Gráfico de dispersão de latitude da espécie *Attila phoenicurus* retirado do site WikiAves.

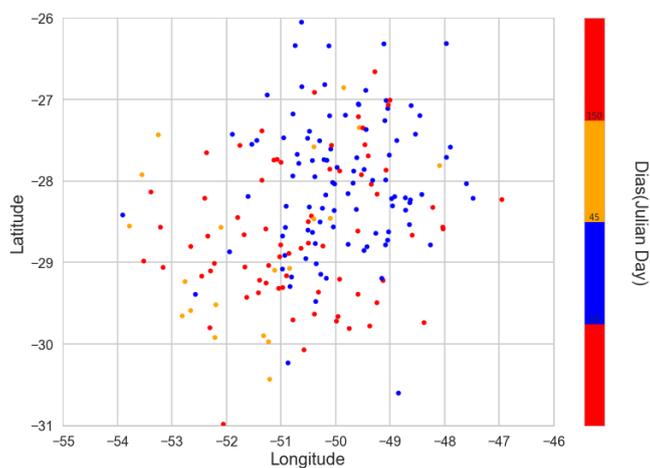


Figura 6. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Amazona pretrei* retirado do site eBird.

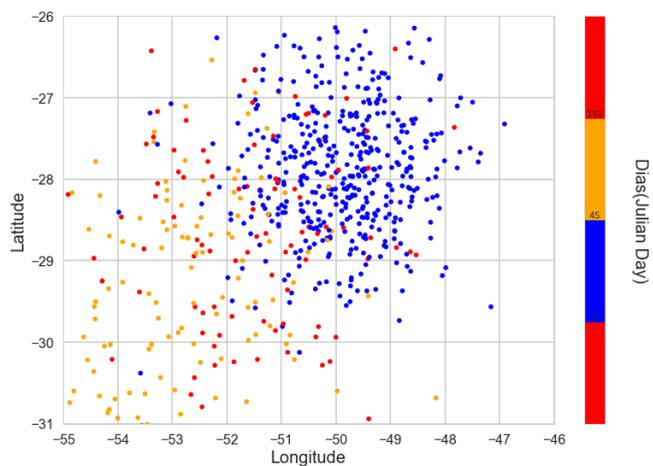


Figura 7. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Amazona pretrei* retirado do site WikiAves.

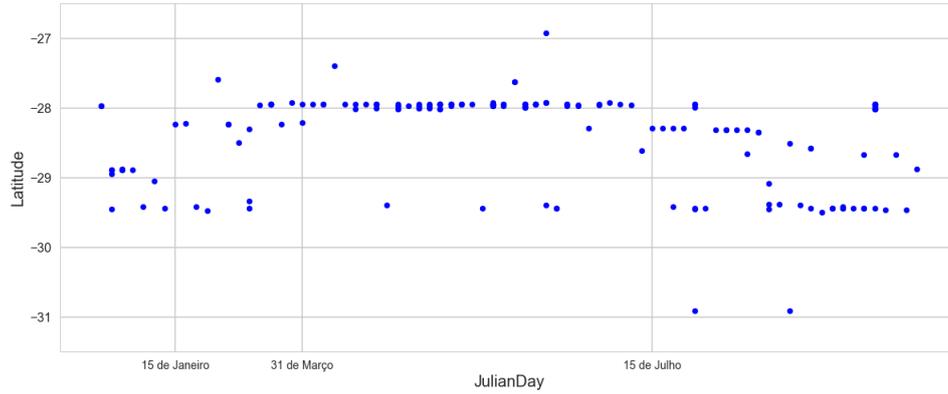


Figura 8. Gráfico de dispersão de latitude da espécie *Amazona pretrei* retirado do site eBird.

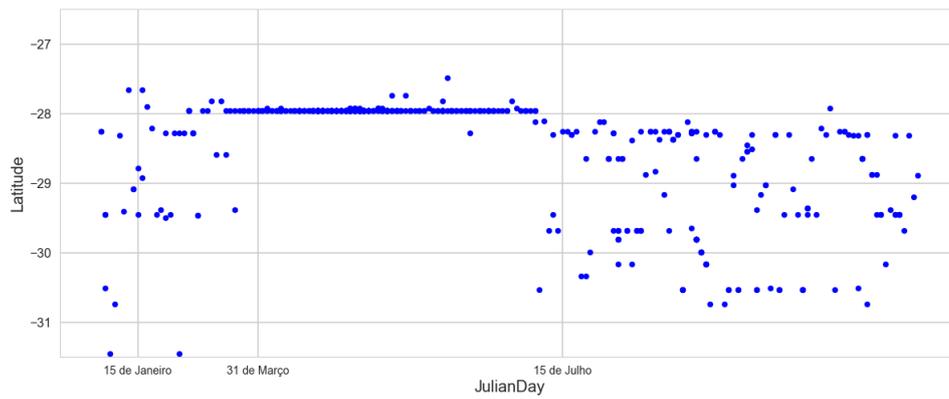


Figura 9. Gráfico de dispersão de latitude da espécie *Amazona pretrei* retirado do site WikiAves.

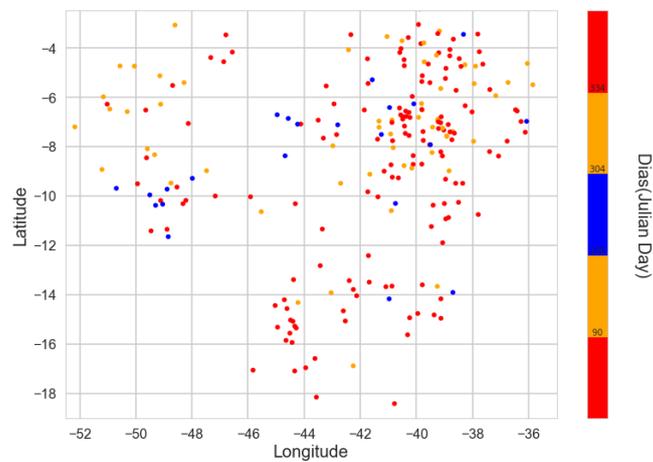


Figura 10. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Casiornis fuscus* retirado do site eBird.

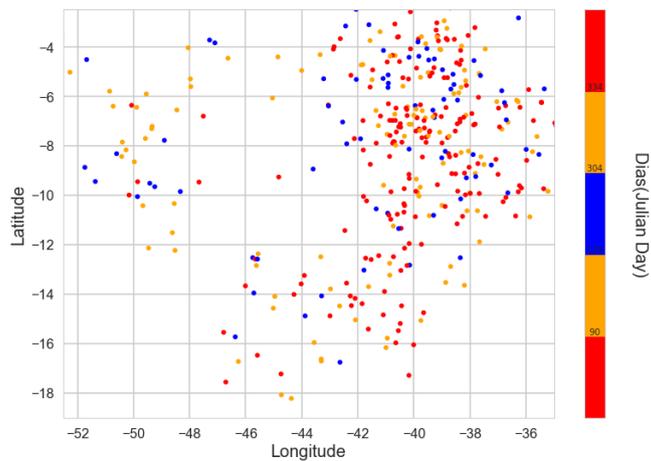


Figura 11. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Casiornis fuscus* retirado do site WikiAves.

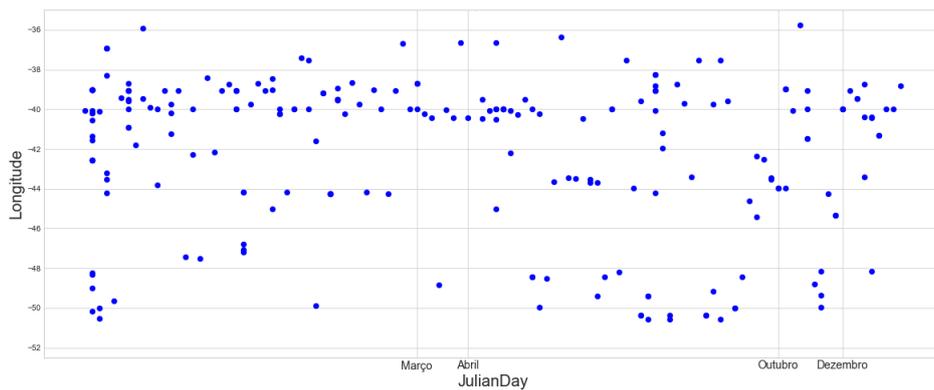


Figura 12. Gráfico de dispersão de longitude da espécie *Casiornis fuscus* retirado do site eBird.

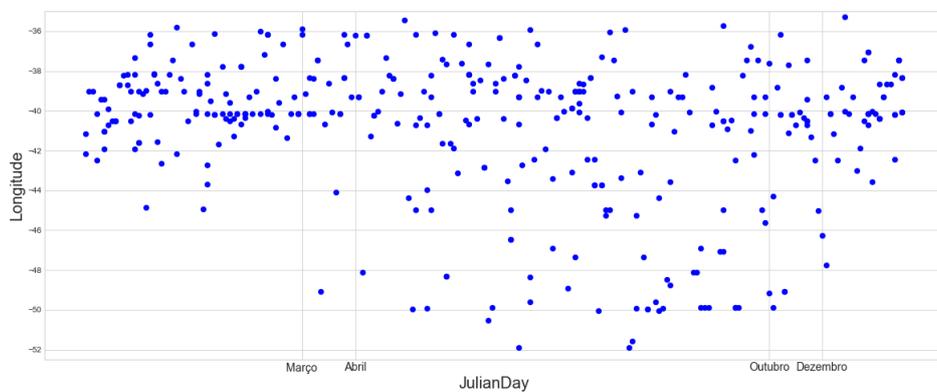


Figura 13. Gráfico de dispersão de longitude da espécie *Casiornis fuscus* retirado do site WikiAves.

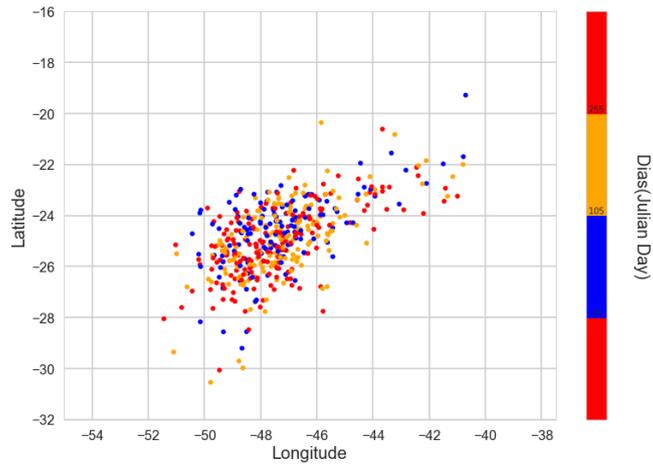


Figura 14. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Tangara peruviana* retirado do site eBird.

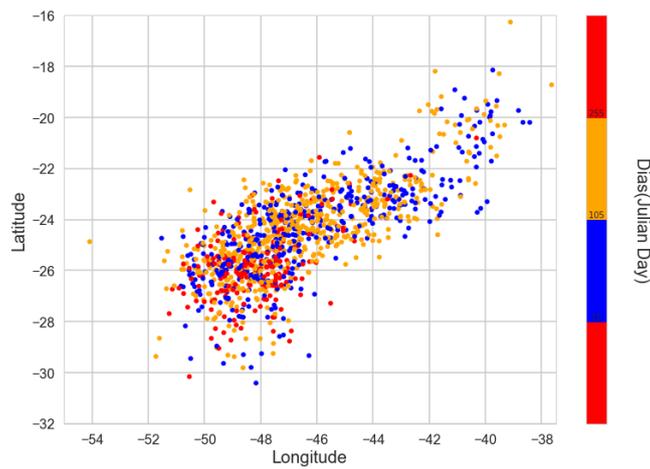


Figura 15. Gráfico de dispersão, com padrão migratório implementado através da colobar, da espécie *Tangara peruviana* retirado do site WikiAves.

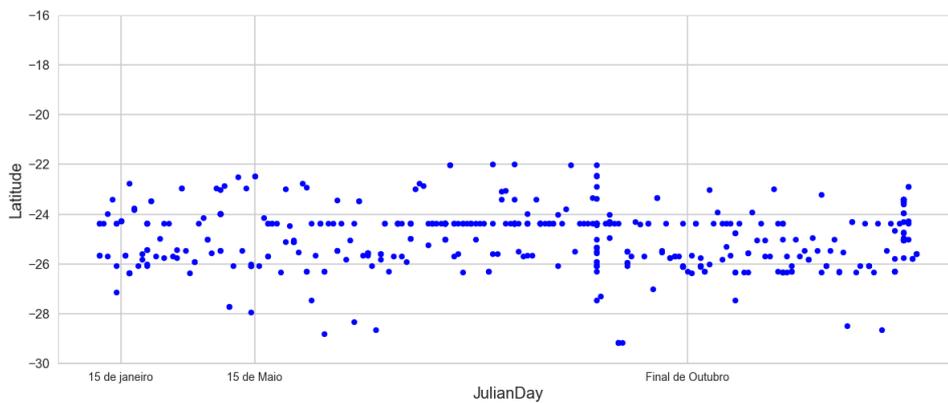


Figura 16. Gráfico de dispersão de latitude da espécie *Tangara peruviana* retirado do site eBird.

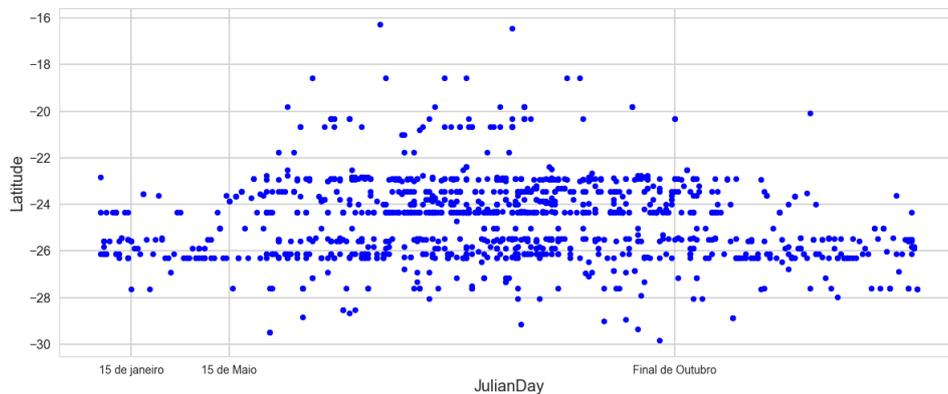


Figura 17. Gráfico de dispersão de latitude da espécie *Tangara peruviana* retirado do site WikiAves.

5. Conclusão e trabalhos futuros

Este trabalho comparou a eficiência e confiabilidade de duas bases de dados diferentes baseadas em registros de Ciência Cidadã, uma Voucherizada e a outra não, através da avaliação dos padrões migratórios de quatro espécies de aves. Essa avaliação foi feita por meio da análise do grande volume de dados obtidos do WikiAves e do eBird e comparados com padrões migratórios pré estabelecidos pela literatura, dessa maneira, pondo à prova a confiabilidade dos registros. Os resultados obtidos mostraram que a base de dados Voucherizada, WikiAves, se mostrou mais confiável revelando padrões migratórios mais condizentes e resultados mais satisfatórios, comparados a base de dados do eBird.

Nós mostramos que a Ciência dos Dados em conjunto com a Ciência Cidadã, que é uma excelente ferramenta para o estudo da migração de aves especialmente se associada com Vouchers, é uma grande ferramenta na confiabilidade dos registros. Isso devido a grande quantidade de dados fornecida nos dias de hoje, facilitando no estudo de padrões de migração de aves.

Como trabalho futuro pretendemos extrair do site do WikiAves, através do Scrapper já implementado, outros campos de dados como os likes de cada foto, o que permitiria filtrar, de maneira automática, eventuais erros de identificação. Outro campo importante é a data em que a foto foi tirada (conforme indicado pela câmera), o que permite detectar erros de lançamento dos dados. Outros campos de interesse, tais como sexo da ave, idade e observações do fotógrafo, serão úteis para novas análises dos dados e facilitarão ainda mais na tomada de decisões.

Referências

- Alves, M. A. S. (2007). Sistemas de migrações de aves em ambientes terrestres no Brasil: exemplos, lacunas e propostas para o avanço do conhecimento. *Revista Brasileira de Ornitologia*, 15(2):231–238.
- Cohn, J. P. (2008). Citizen science: Can volunteers do real research? *AIBS Bulletin*, 58(3):192–197.

- eBird (2012). ebird. 2012. ebird: An online database of bird distribution and abundance [web application]. ebird, ithaca, new york. disponível em: <http://www.ebird.org>. acesso em: 10 mar. 2018.
- Lees, A. C. and Martin, R. W. (2015). Exposing hidden endemism in a neotropical forest raptor using citizen science. *Ibis*, 157(1):103–114.
- Madhavan, S. (2015). *Mastering Python for Data Science*. Packt Publishing Ltd.
- Martinez, J. and Prestes, N. (2002). Ecologia e conservação do papagaio-charão amazona pretrei. *Ecologia e conservação de psitacídeos no Brasil*, pages 173–192.
- Ramesh, V., Gopalakrishna, T., Barve, S., and Melnick, D. J. (2017). Iucn greatly underestimates threat levels of endemic birds in the western ghats. *Biological Conservation*, 210:205 – 221.
- Resnik, D., Elliott, K., and K. Miller, A. (2015). A framework for addressing ethical issues in citizen science.
- SiBBr. Sistema de informação sobre a biodiversidade brasileira (sibbr). disponível em: <http://www.sibbr.gov.br/cienciacidada/#/about>. acesso em: 12 mai. 2018.
- SiBBr(1). Projeto wikiaves, disponível no site do sibbr. disponível em: <http://www.sibbr.gov.br/cienciacidada/#/portfolio/projetos/wikiaves/>. acesso em: 12 mai. 2018.
- Somenzari, M., Amaral, P. P. d., Cueto, V. R., Guaraldo, A. d. C., Jahn, A. E., Lima, D. M., Lima, P. C., Lugarini, C., Machado, C. G., Martinez, J., et al. (2018). An overview of migratory birds in brazil. *Papéis Avulsos de Zoologia*, 58.
- Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., et al. (2014). The ebird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169:31–40.
- Turnhout, E., Lawrence, A., and Turnhout, S. (2016). Citizen science networks in natural history and the collective validation of biodiversity data. *Conservation Biology*, 30(3):532–539.
- Waller, M. A. and Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2):77–84.
- WikiAves.com (2008). Wikiaves. disponível em: <http://www.wikiaves.com.br/>. acesso em: 02 jan. 2018.