

Exploratory Data Analysis

Sample Size

Ao todo foram capturados 22 tucos. Desse número nem todos receberam acelerômetro e lightlogger. As tabelas abaixo mostram quantos animais foram capturados por campanha e quantos animais receberam acelerômetro ou lightloggers.

Table 1: Número de tucos capturados por campanha.

Month	n	Accelerometer	Lightlogger
February	5	5	5
July	8	8	6
March	2	2	0
October	7	7	2
TOTAL	22	22	13

Table 2: Número de tucos capturados por campanha separados por sexo.

season	sex	n	Accelerometer	Lightlogger
February	f	3	3	3
February	m	2	2	2
July	f	5	5	4
July	m	3	3	2
March	f	2	2	0
October	f	5	5	2
October	m	2	2	0

Data Pre-processing

Os dados de acelerômetro são coletados em 10Hz. Para transformarmos os dados da aceleração de cada eixo em um índice de atividade calculamos o VeDBA. Esse índice é a soma vetorial dos valores de aceleração dos valores de cada eixo descontado o valor da gravidade. Depois de calculado o VeDBA vamos resumir os dados para tornar as análises mais convenientes computacionalmente e também melhorar a visualização desses dados. Os dados de VeDBA foram resumidos a cada um minuto usando-se a mediana dessa janela de tempo, o que corresponde a uma mediana móvel sem sobreposição a cada 600 amostras. A maneira que podemos *interpretar esses dados* é que cada amostra na verdade representa o valor de VeDBA que o animal teve por pelo menos 30s consecutivos. Ou seja, caso uma amostra tenha valor de 0.4 de VeDBA podemos afirmar que naquele minuto o animal se manteve com um VeDBA maior ou igual à 0.4 durante 30 segundos consecutivos. Isso talvez tenha mais importância depois que classificarmos as amostras usando o HMM.

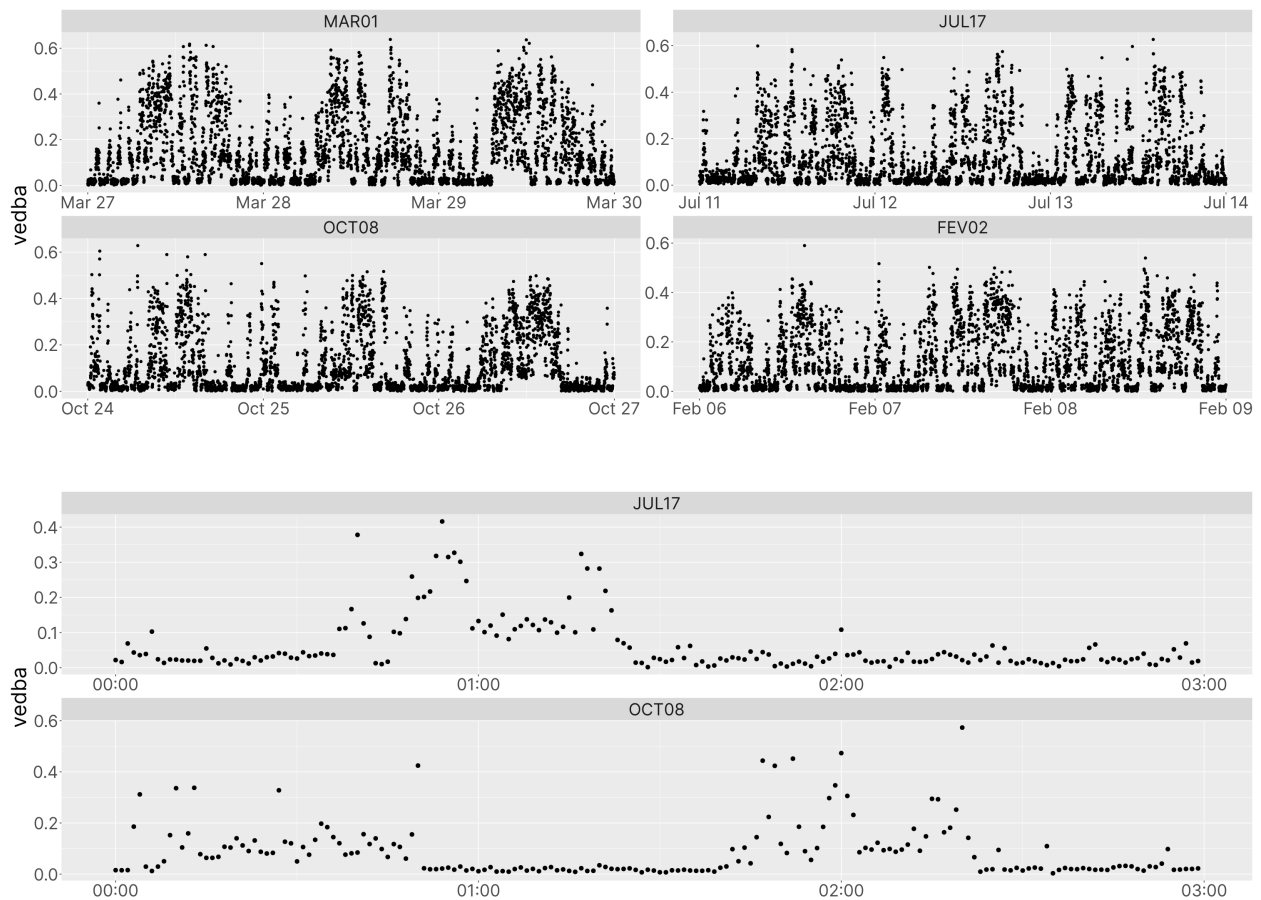
Os dados de exposição à luz são amostrados a cada 1 minutos mas só registram a amostra com maior valor

a cada 5 minutos. Para esses dados o que penso ser a melhor saída é estabelecer um limiar e considerar que qualquer valor maior que esse limiar é um episódio de exposição à luz. Para isso sigo usando o limiar de 2 lux [Milene]. Importante também pensarmos que temos dados de acelerômetro e lightlogger que são desiguais temporalmente e talvez seja necessário nos atentarmos pra isso quando sobrepusermos esses dados.

Time Series Plot

Primeiro uma verificada geral nos dados em formato de série temporal. Os gráficos mostram apenas os 4 primeiros dias de registro de cada animal.

Não é possível ver muita diferença entre os animais então nas séries temporais estão plotados apenas 4 animais. É interessante que em alguns animais já é possível ver que os pontos parecem estar mais ou menos organizados em três regiões distintas na vertical.

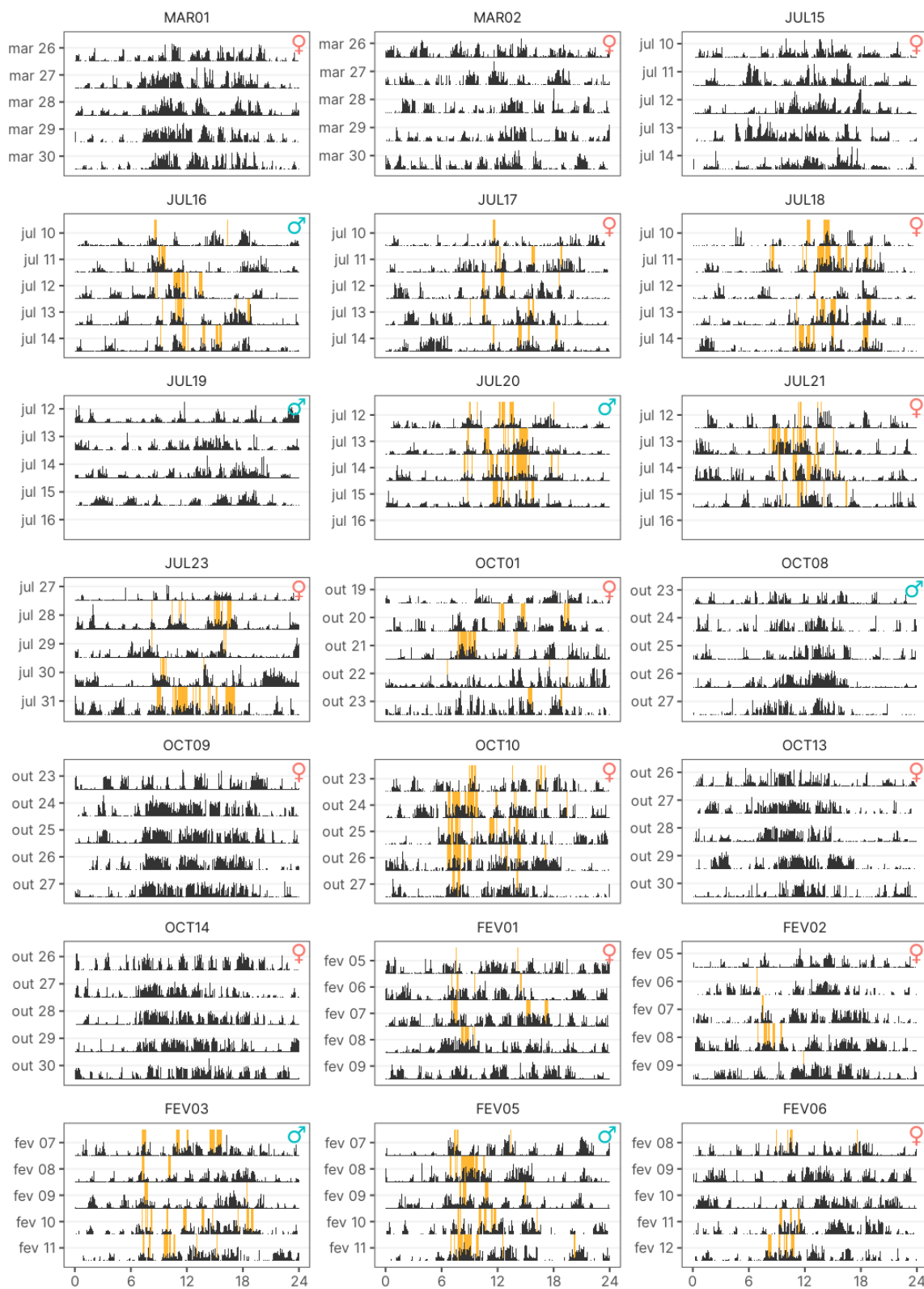


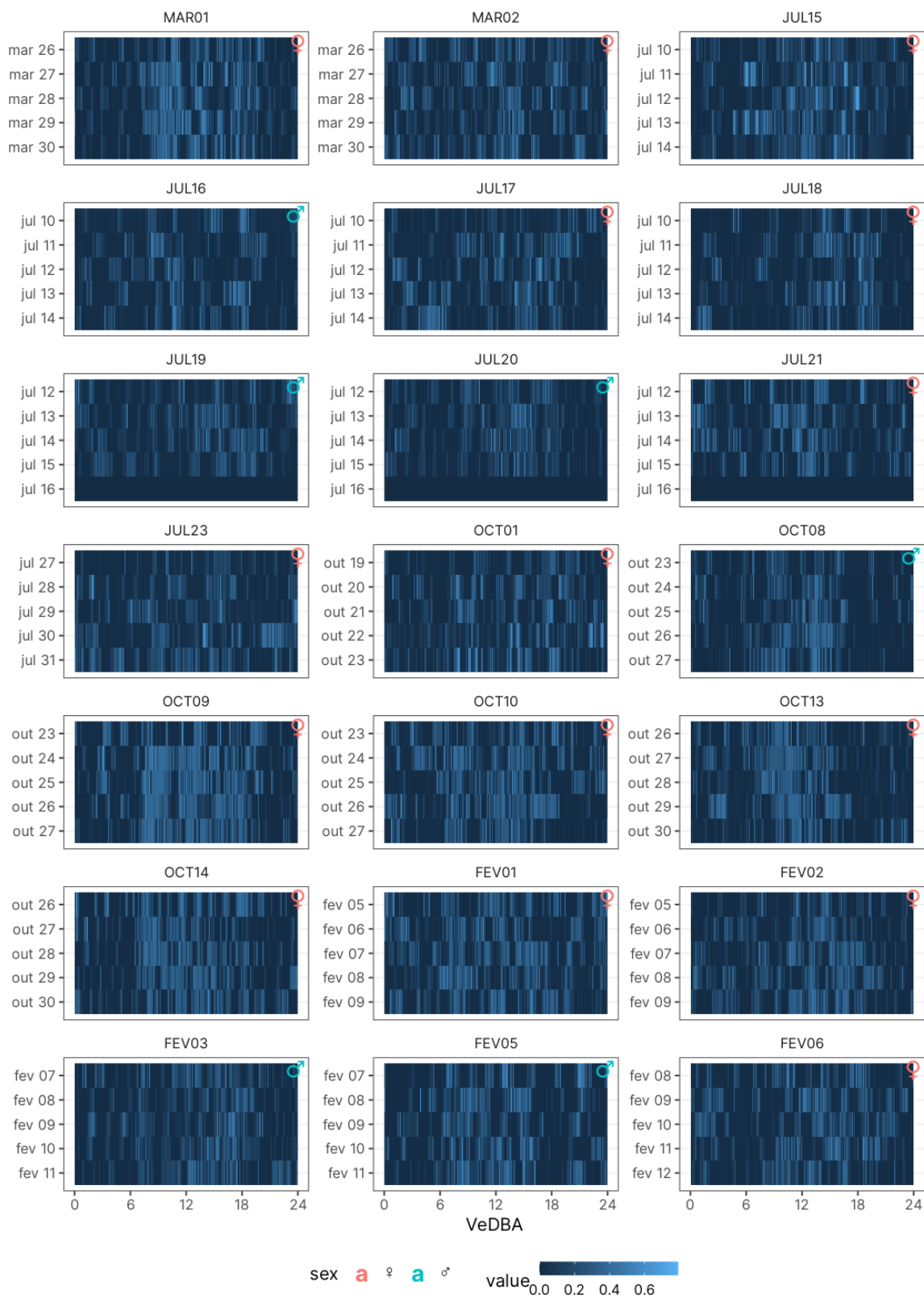
Actogramas

Os actogramas estão divididos por animais, com o sexo no topo direito. Os dados de acelerômetro estão em preto, sem nenhum limiar aplicado aos gráficos. Os dados de exposição à luz estão em laranja, todos registros maiores do que 2 lux são considerados como uma saída e plotados como barras nos gráficos.

O segundo conjunto de actogramas estão num formato de heatmap. Esses actogramas possuem apenas os dados de atividade e foram plotados apenas para ver se observávamos melhor algum padrão em relação aos actogramas de barras.

O que é possível observar nos actogramas, e que não é muita surpresa, é que parece realmente existir algum ritmo circadiano nos animais de campo. Esse ritmo é mais marcado em alguns animais do que em outros, independente da época de captura, como por exemplo entre os animais MAR01 e MAR02.





Histogramas e Gráficos de Densidade

Antes de seguir para outras análises vamos observar as distribuições dos dados de acelerômetros para ver se há muita heterogeneidade entre os animais. Os dados de atividade para esses gráficos foram limitados à 4 dias por animal, assim o número de amostras por animal é o mesmo.

As distribuições de VeDBA parecem ter um range muito próximo de valores. Ou seja, parece não haver animais que possuam uma atividade muito mais intensa do que outros, o que também pode ser visto na tabela abaixo. Porém, o formato da distribuição muda entre alguns animais, principalmente entre estações. Por exemplo, os animais capturados em outubro parecem ter maior número de amostras com valores mais à direita da distribuição, entre 0.2 e 0.5. Alguns animais capturados em Fevereiro parecem continuar com essa tendência. Em julho, porém, os animais parecem ter uma distribuição com maior concentração em valores mais centrais, entre 0.05 e 0.2. Essas mesmas tendências também podem ser observadas de uma forma mais compacta no gráficos de densidade por animal.

Então, os animais, apesar de terem distribuição de VeDBA que não distoam muito uns dos outros no seu range, parecem passar tempos diferentes em tipos de diferentes de comportamentos.

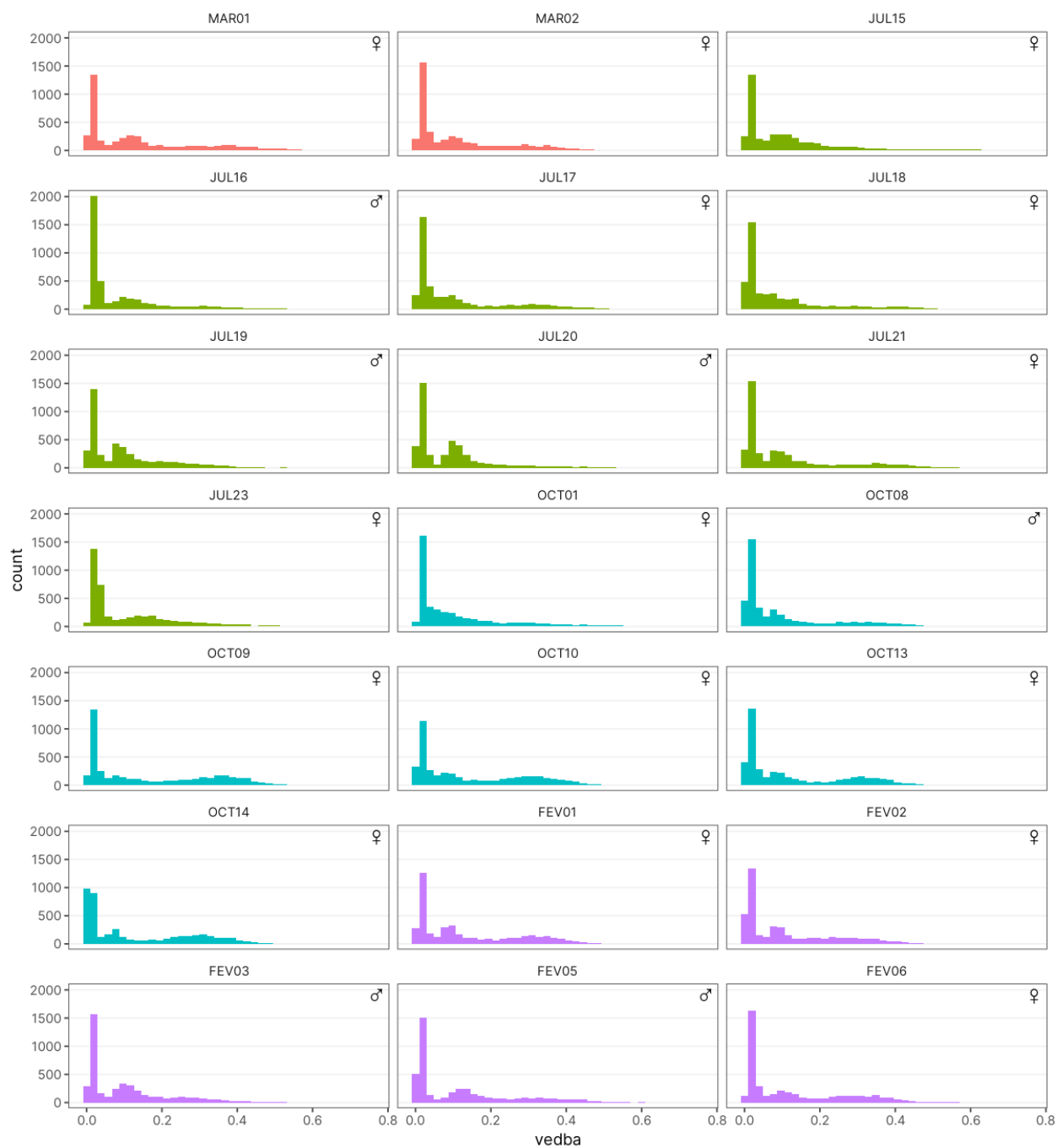


Figure 1: Histograma dos valore de VeDBA por animal (cada quadro). As cores representam os meses em que esses animais foram capturados.

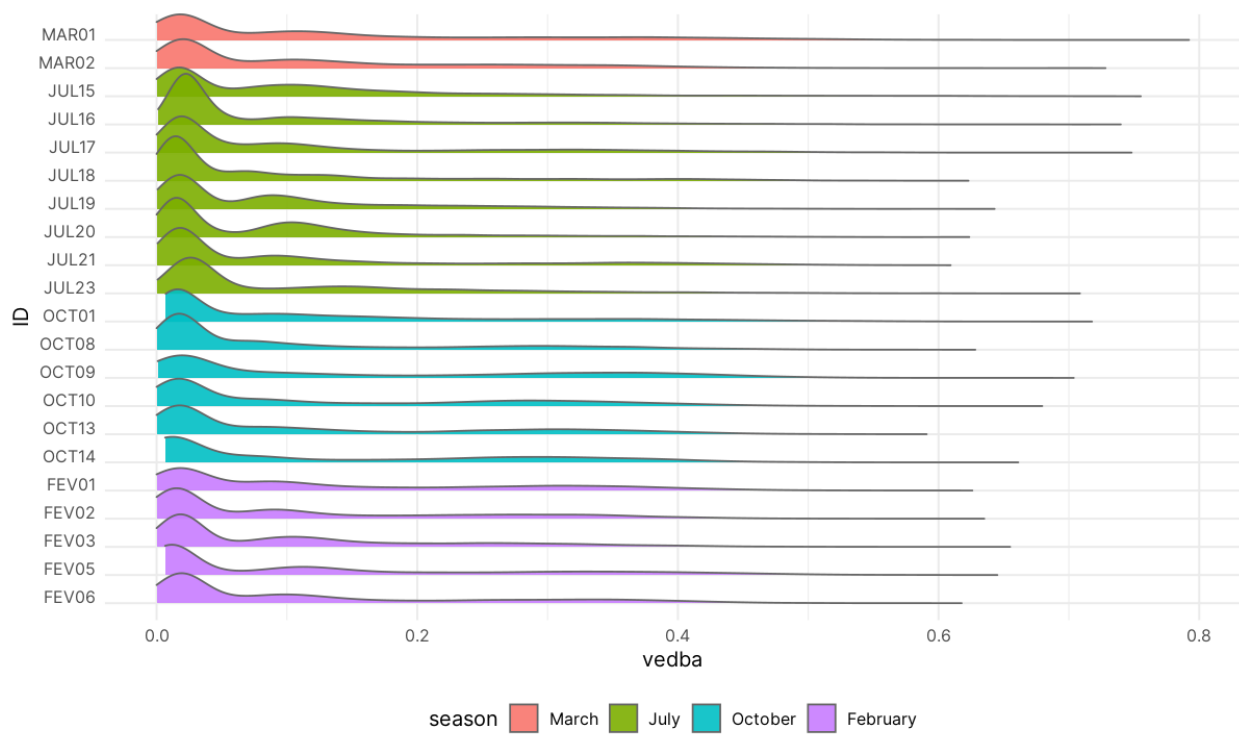
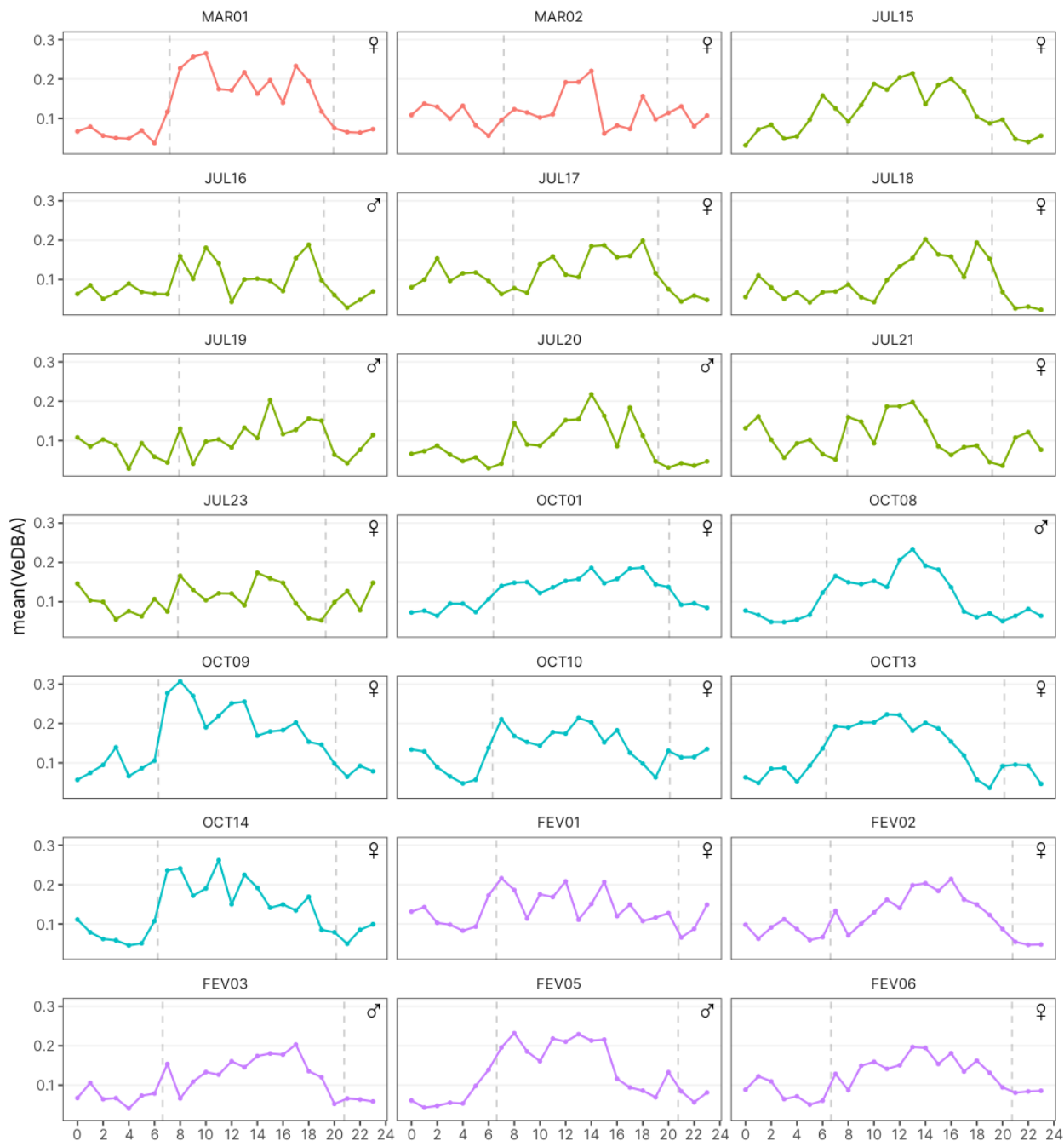


Figure 2: Ridge plot mostrando a densidade da distribuição dos valores de VeDBA divididos por animal. As cores representam os meses em que os animais foram capturados.

Padrões Médios de Atividade por Animal

Para terminar, é interessante ver como é o “ritmo médio” ao longo de todo tempo de registro dos animais. Acho que esse gráfico é especialmente interessante para pessoas fora da cronobiologia, já que é bem mais intuitivo do que os actogramas.

Nesses primeiro gráfico foi plotado a média por hora dos valores de VeDBA de cada animal.



Gráficos Agrupados por Estação e Sexo

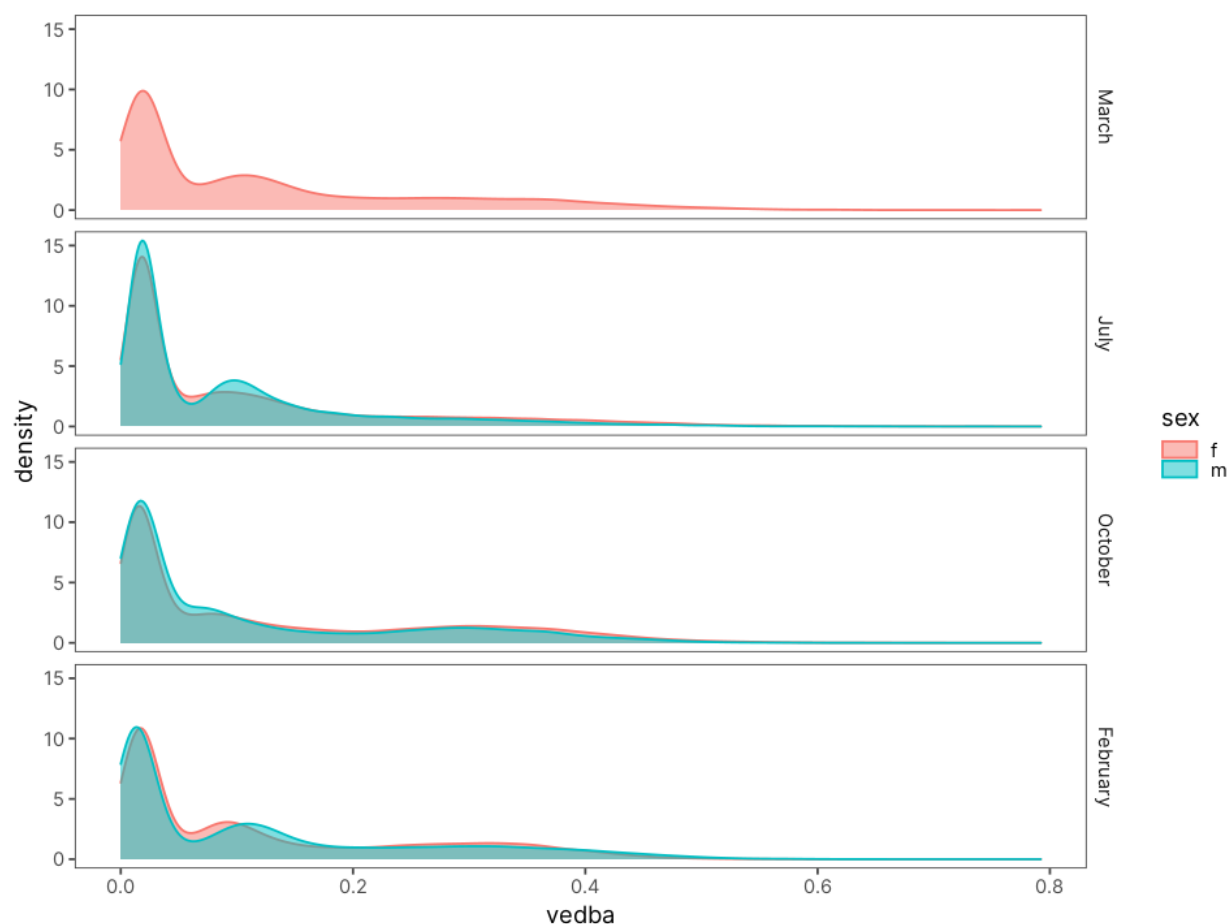
Histogramas/Gráfico de Densidade

Visualmente parece haver alguma diferença sazonal nas atividade dos tucos. Vamos inspecionar melhor isso agrupando os dados por estação e sexo.

Como o número de amostras varia entre estações e sexo o histograma não seria muito informativo nesse caso. Então pra isso podemos fazer um gráfico de densidade. Esses gráficos basicamente refletem a forma do histograma mas, nesse caso, tem a vantagem de que são padronizados para que a área abaixo da curva seja sempre igual a 1.

Assim, podemos ver como o perfil das distribuições muda ao longo das estações e entre sexos mesmo tendo um número de amostras diferente para cada condição. Os valores em y não significam muita coisa individualmente. O mais importante, nesse caso, é que as áreas de todas as curvas são iguais. Para uma discussão interessante sobre a diferença entre histogramas e gráficos de densidade probabilística: <https://towardsdatascience.com/histograms-and-density-plots-in-python-f6bda88f5ac0>.

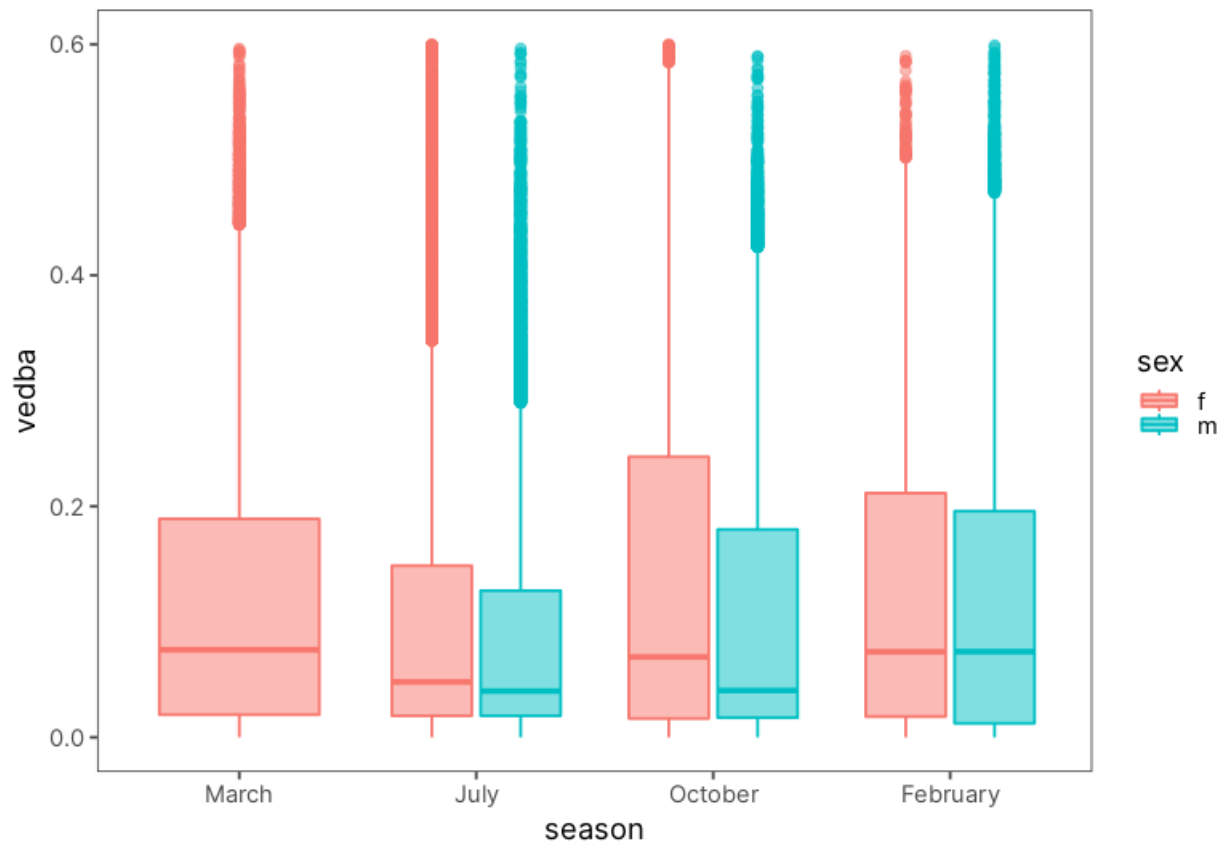
Visualmente temos o que também vimos nos histogramas anteriores. A distribuição parece mudar ao longo do ano. No entanto, quando sobrepomos as distribuições de cada sexo parece não haver muita diferença entre machos e fêmeas.



Medianas e Boxplots

season	sex	median	mean	n.animals	n.points
March	f	0.0763851	0.1241451	2	17280
July	f	0.0485506	0.1081949	5	41760
July	m	0.0402647	0.0933556	3	18720
October	f	0.0697453	0.1326020	5	74880
October	m	0.0406470	0.1104647	1	21600
February	f	0.0740544	0.1220490	3	33120
February	m	0.0742569	0.1214579	2	28800

Warning: Removed 213 rows containing non-finite values (stat_boxplot).



Padrão Médio de Atividade por Estação e Sexo

Nesse último gráfico eu queria verificar se temporalmente existe alguma diferença entre machos e fêmeas. Cada painel é um grupo entre as opções de combinação de sexo e estação.

As mesma curvas de VeDBA médio por hora são mostradas em cinza ao fundo, onde cada curva representa um animal diferente. Em cores, colorido por estação, está a média de VeDBA por hora do grupo representado no painel.

Aparentemente também não existe uma diferença tão grande entre machos e fêmeas quanto aos horário de atividade. Porém, temos um n bem baixo de machos capturados o que torna difícil fazermos uma observação muito concreta.

