**Analysis of Armadillo Acceleration Data**

November 9, 2020

Armadillo movement data were analyzed in multiple ways to determine if “resting”, “foraging”, and “transit” behaviors could be classified. Armadillo acceleration (measured as activity counts per recording interval; 5 min) and GPS data (recorded at 7 min intervals) were recorded and were analyzed in bayesmove using the segmentation and clustering (LDA) procedures.

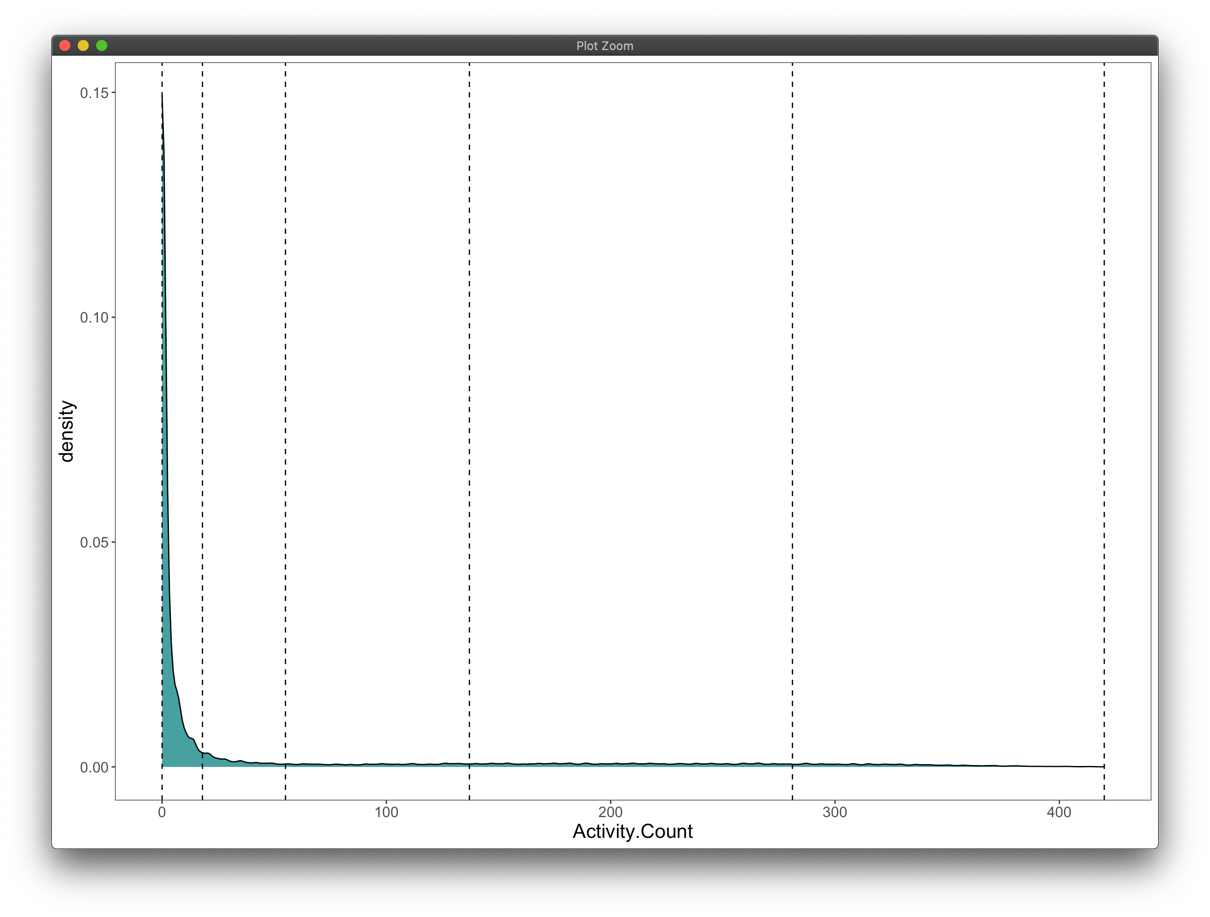
*Data preparation*

Since there were numerous observations of missing data for Activity Count and GPS data, these data were first filtered to remove all duplicates and only include Activity Counts that were ±1 min from the 7 min time step for the GPS tag (since this was the data stream treated as a reference). Data were filtered in this way since we have spatial movement patterns from the GPS locations, which were used to calculate Step Lengths and Turning Angles. This left 67,873 observations from 7 individuals for analysis by the model. Within this dataset, there was 2.6× more Activity Count data than GPS data (step lengths, turning angles).

Since Activity Count and Step Length were highly right-skewed, quantiles were used to bin these variable in a way that was biologically relevant. Specifically, I discretized Activity Count and Step Length into 5 bins each. Activity Count was discretized using the 0, 0.75, 0.80, 0.85, 0.95, and 1 quantiles, whose bin limits were 0, 18, 55, 36, 137, 281, and 420 counts of deviation from the previous values of the accelerometer. Step Length was discretized using the 0, 0.50, 0.75, 0.95, 0.99, and 1 quantiles, whose bin limits were 0, 33, 101, 357, 931, and 1957 m for a 7 min time step. Turning Angle distributions were quite flat for most individuals (except ‘Mazeboti’), so equal bin widths of was used on .

Chart, scatter chart

Description automatically generated



A picture containing chart

Description automatically generated

A picture containing water, boat, table, sitting

Description automatically generated

Figure 3 This turning angle distribution for all armadillos is biased by a single individual (‘mazeboti’), that possess one of the largest sample sizes and includes a high density of values near 0 radians.

*Segmentation model*

Each unique day was used to pre-specify breakpoints for the model per ID since there was a clear temporal pattern of activity each day. Based on traceplots of the log marginal likelihood for each individual, the model appeared to converge after 20000 iterations.

A picture containing building, window

Description automatically generated

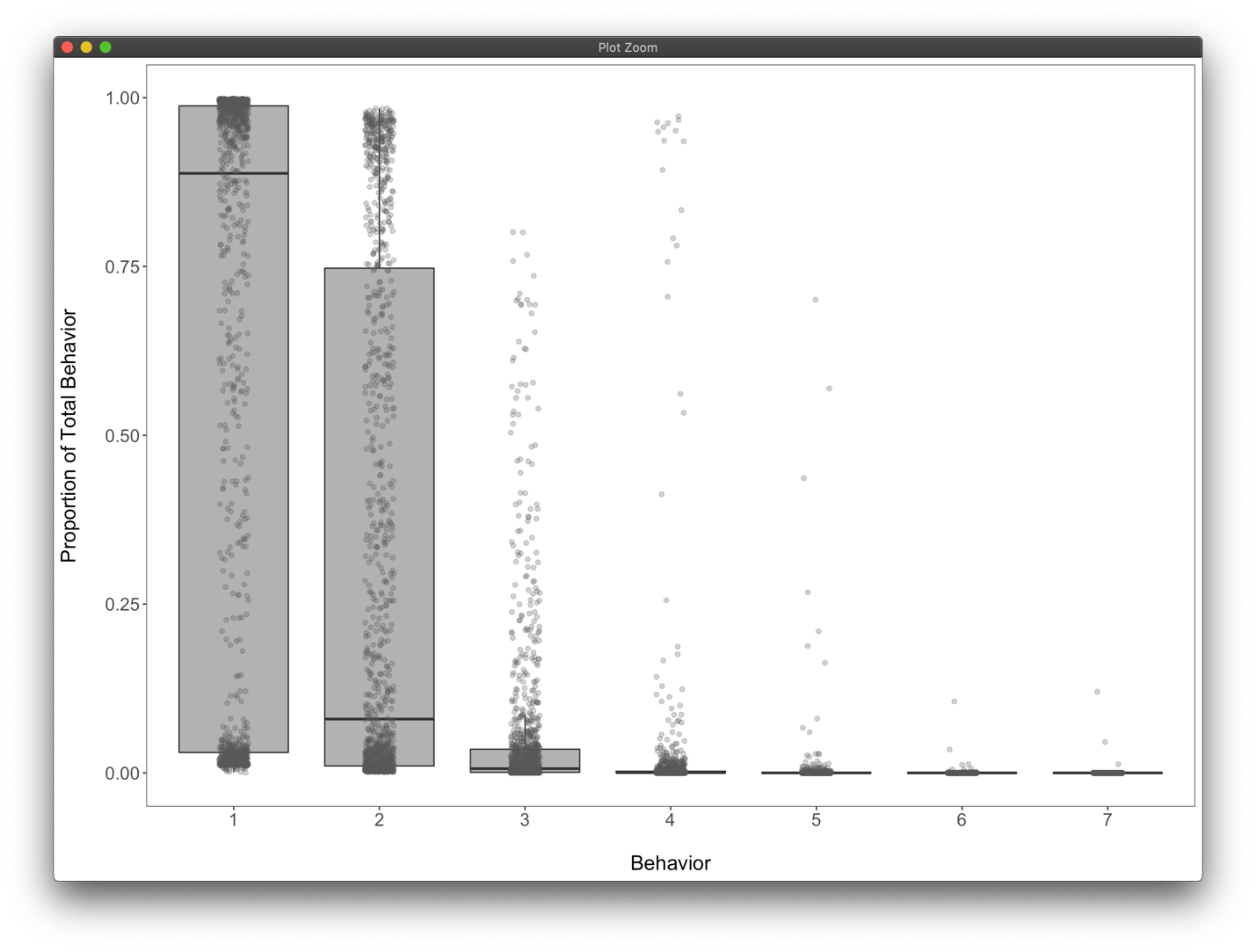
A close up of a window

Description automatically generated

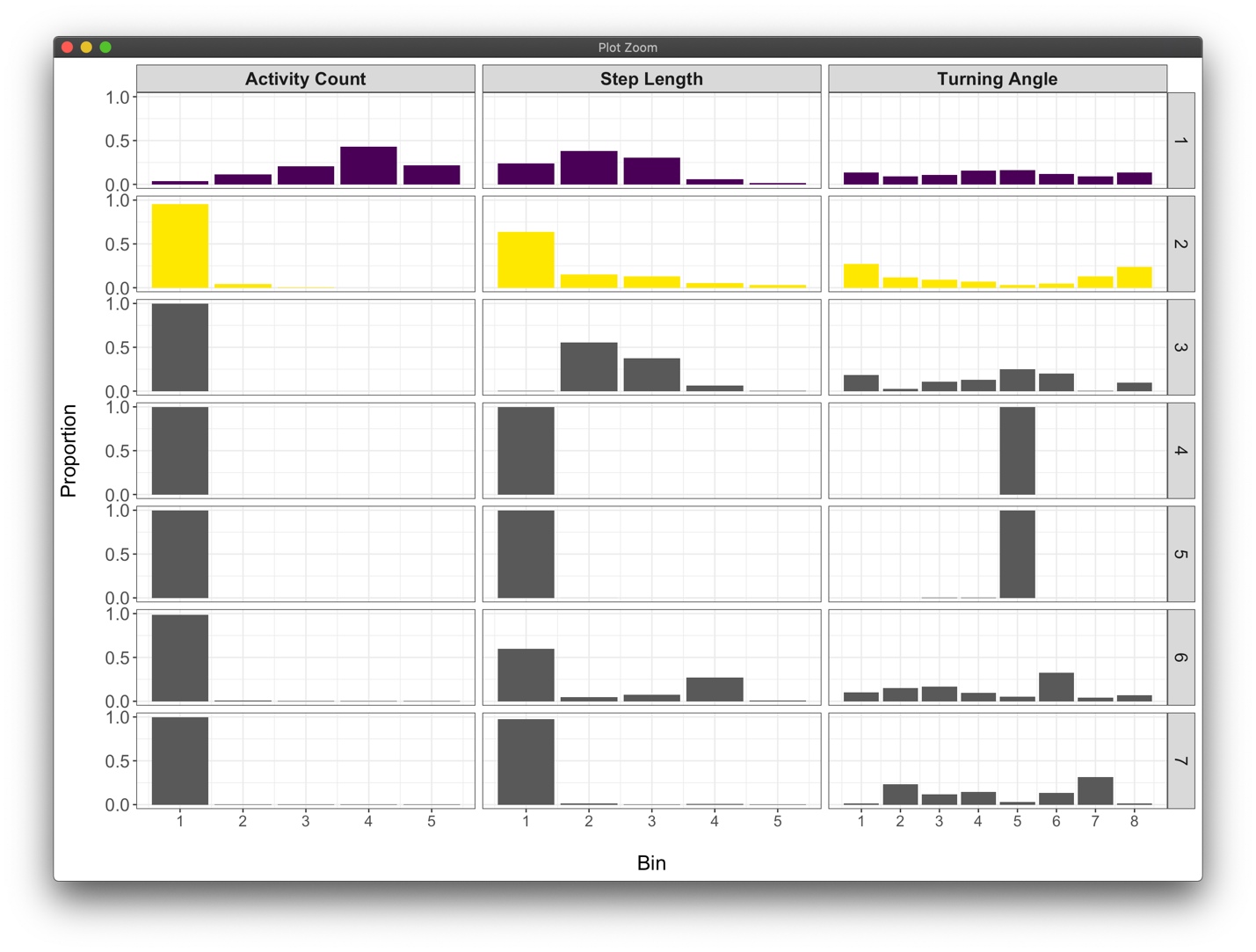
Only a couple examples of the segmentation results are shown to demonstrate breakpoint estimation. Most IDs (e.g., ‘gala’) estimated breakpoints at least once a day if not more frequently. However, ‘mazeboti’ has 3 segments in the middle of its time series that are each quite long (i.e., comprised of multiple days/weeks). This appears to be due to the lack of available Activity Count data during these segments, but that still has yet to be confirmed. These data did not include time spent resting in the burrow during the day, so results presumably reflect active behavioral states only. Additionally, thousands of observations of Activity Count data (w/o corresponding GPS data) were omitted from these analyses since they overwhelm the step lengths and turning angles that were analyzed.

*LDA model*

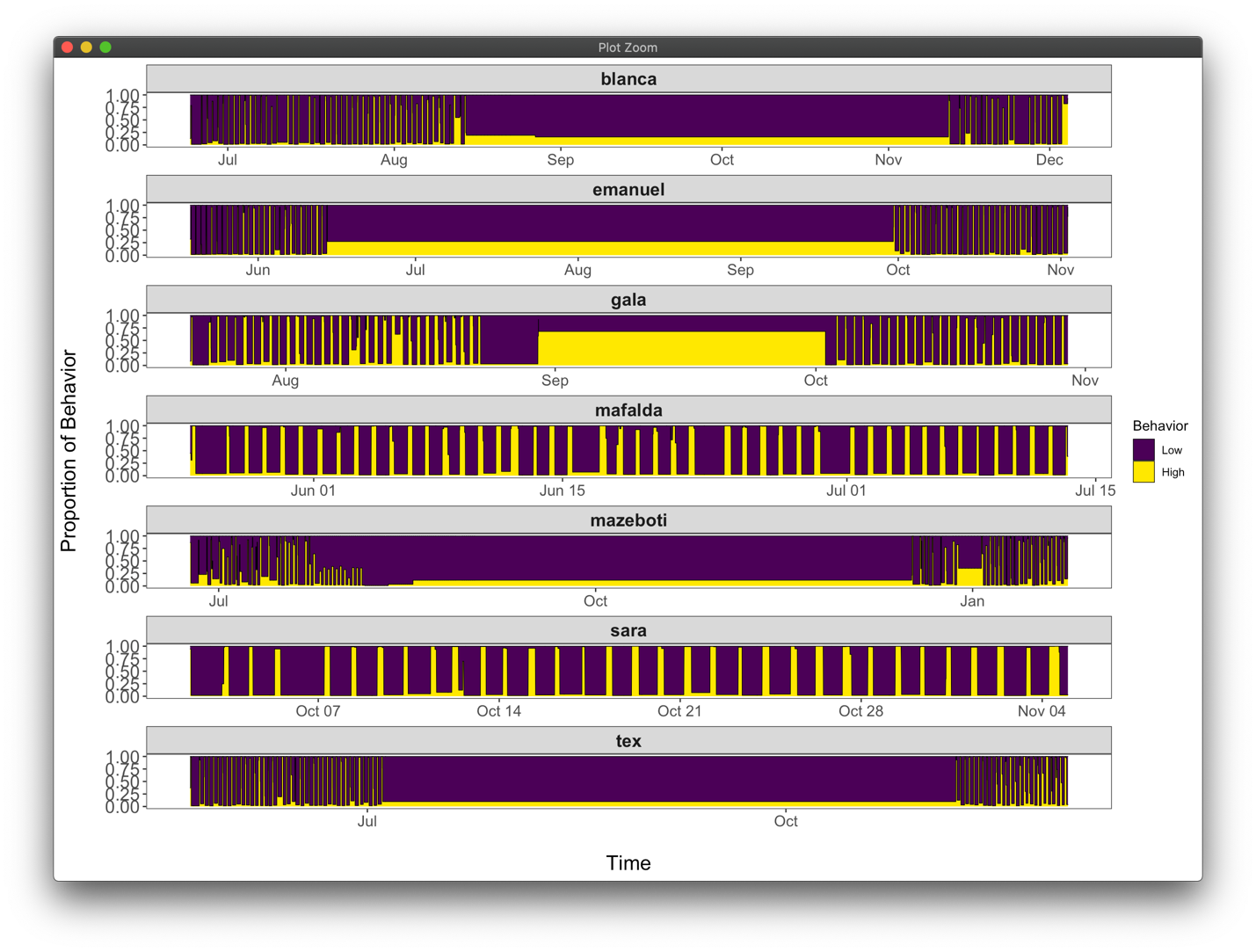
Following the clustering of the segments across all 7 individuals, plots were inspected to determine the optimal number of behavioral states as well as the state-dependent distributions.



Based on this boxplot of results from the theta matrix, the first 2 states/clusters accounted for 93.8% of all observations on average. Therefore, these clusters were retained for further analysis.



Plots of the bin distributions for each of the 7 possible states indicate that these first 2 states are biologically distinct and relevant. Based on these distributions, state 1 appears to represent a high activity “transit” state, whereas state 2 represents a low activity “foraging” state. It should be mentioned that presumed “burrow” behavior was not identified by the model since GPS data were not available during the day (and Activity Counts were only kept when matching GPS observations).



Time series plots of these behavior proportions for each individual armadillo are shown next and clearly depict regular intervals of higher and lower activity. The long periods in the middle of the time series with minimal change in behavior proportion is likely an artefact of the visualization function used, meaning that there are large gaps of missing data during these periods.

Graphical user interface, website

Description automatically generated

*Conclusions*

Based on these findings, it appears that our model performs well at characterizing latent states from a combination of Activity Counts, Step Length, and Turning Angle data streams. While only this single analysis is included here, previous analyses that only evaluated Step Length and Turning Angle or that tried to generate fake Step Lengths and Turning Angles during “burrow” behavior resulted in breakpoints that did not identify daily patterns and therefore improperly classified behavioral states. Therefore, “burrow” behavior may need to be classified for observations outside of the model.

The next potential step, if wanting to classify all Activity Count data, would be to make predictions based on our model results. This would depend on how closely the Activity Count distributions match from the “held out” data compared to those that were classified.