Data analysis and visualizations of drosophila behavioral phases

Xiaona Zhou
CUNY New York City College of Technology

Data analysis and visualizations of drosophila behavioral phases

Xiaona Zhou[1], Professor Boyan Kostadinov[1]

[1] Department of Mathematics, New York City College of Technology, CUNY

Abstract

We analyze drosophila behavioral data from the neuroscience labs of Professor Maria de la Paz Fernandez, Barnard College of Columbia University, and Professor Orie Shafer, the Neuroscience Initiative of the CUNY Advanced Science Research Center. The main goal of this project is to create algorithms for analyzing and visualizing the average activity of drosophila across a specified number of days and across all live flies, and use this analysis to calibrate a smoothing filter to be applied to the raw fly activity so that the drosophila behavioral phases can be computed and visualized. We also investigate how to compute and visualize the onset and offset of behavioral phases as well as the trend lines of mid-day fly activity. We developed a **Shiny App** that implements all data analysis and visualizations.

Introduction

The data we used in this project are derived from the activity of fruit flies from several experiments as part of the research conducted in (Fernandez et al. 2020) and (Kostadinov et al. 2020). In this poster, we discuss how the average activity plot is obtained, how to compute the evening phases, which are related to the peak of activity closest to the lights-off time point in the experiments, and how to compute the trendlines of mid-day peaks of fly activity. The main results covered in this poster are the following:

1. Analyze and visualize average activity of drosophila across a specified number of days and across all live flies.
2. Apply smoothing filter to the raw fly activity so that the drosophila behavioral phases can be computed.
3. Compute and visualize the onset and offset of behavioral phases as well as the trend lines of mid-day fly activity.
4. Build a **Shiny App** implementing all data analysis.

Average Activity Plot

After data pre-processing, we remove all dead flies from the data. A fly is considered dead when there is zero activity. We average bin-wise the fly activity across all live flies and across the specified number of days. We visualize the average activity plot in Figure 1.

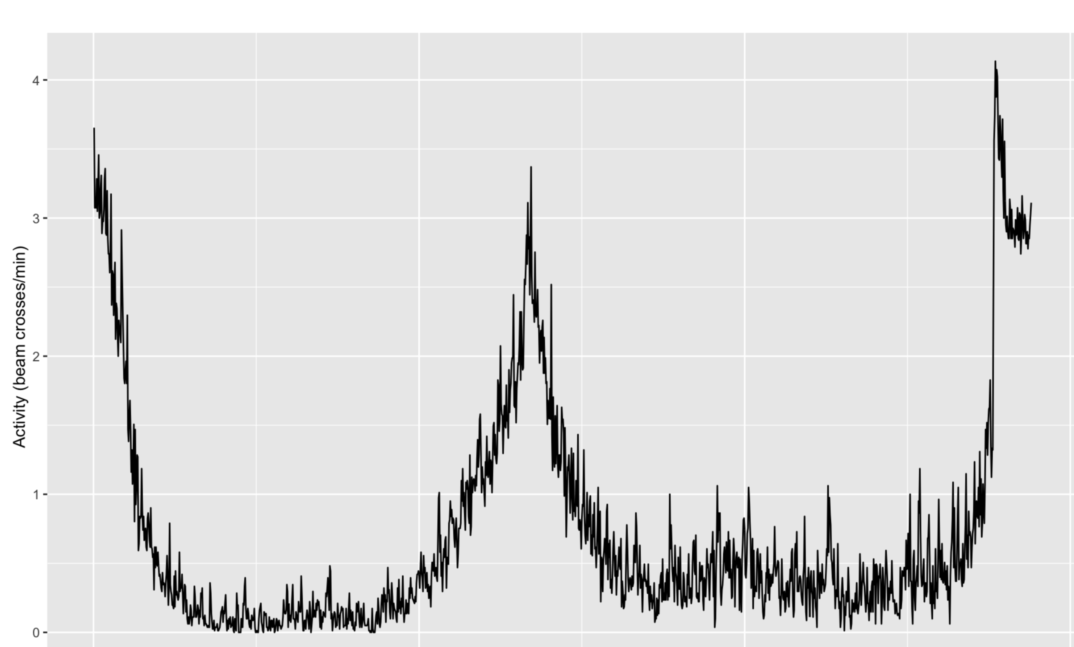

Figure 1: Average activity plot over three days

Phases

An evening phase of a fly is the time difference between the peak of fly activity, closest to lights-off, and the lights-off time point itself, so the phases in units of time could be positive or negative. First, we use the average activity plot to calibrate a Butterworth filter to smooth out the raw activity data, and then we find the peak closest to lights-off using the `findpeaks` function from the **pracma** package. The algorithm behind *findpeaks* returns a matrix in the form shown below:

$$\begin{bmatrix} \text{peak1 height} & \text{peak1 index} & \text{peak1 start} & \text{peak1 end} \\ \text{peak2 height} & \text{peak2 index} & \text{peak2 start} & \text{peak2 end} \\ . & . & . & . \end{bmatrix}$$

The second column gives the location of the peak, the third and forth columns give the indices where the peak pattern begins and ends. The algorithm is summarized into the four steps listed below:

1. Convert the data into “+”, “-”, “0”. For example, if a number is greater than the previous one, then the result is a “+”, if smaller then “-”, and “0” if the same.
2. Find the pattern: the pattern consists of n “+” followed by m “-”. Note that “0” terminates the pattern, and a new pattern must start with “+”.
3. Find the maximum peak within each pattern and get its value and index.
4. Check whether the peaks found are “valid” using three criteria:
   - Is peak value > minimum peak height?
   - Is the difference between peak’s height and lowest height within the pattern greater than the threshold? This eliminates small peaks.
   - A peak is not valid when it is too close to a bigger peak.

An evening phase plot is shown in Figure 2, where the shaded area represents “night”, after the lights-off time point.

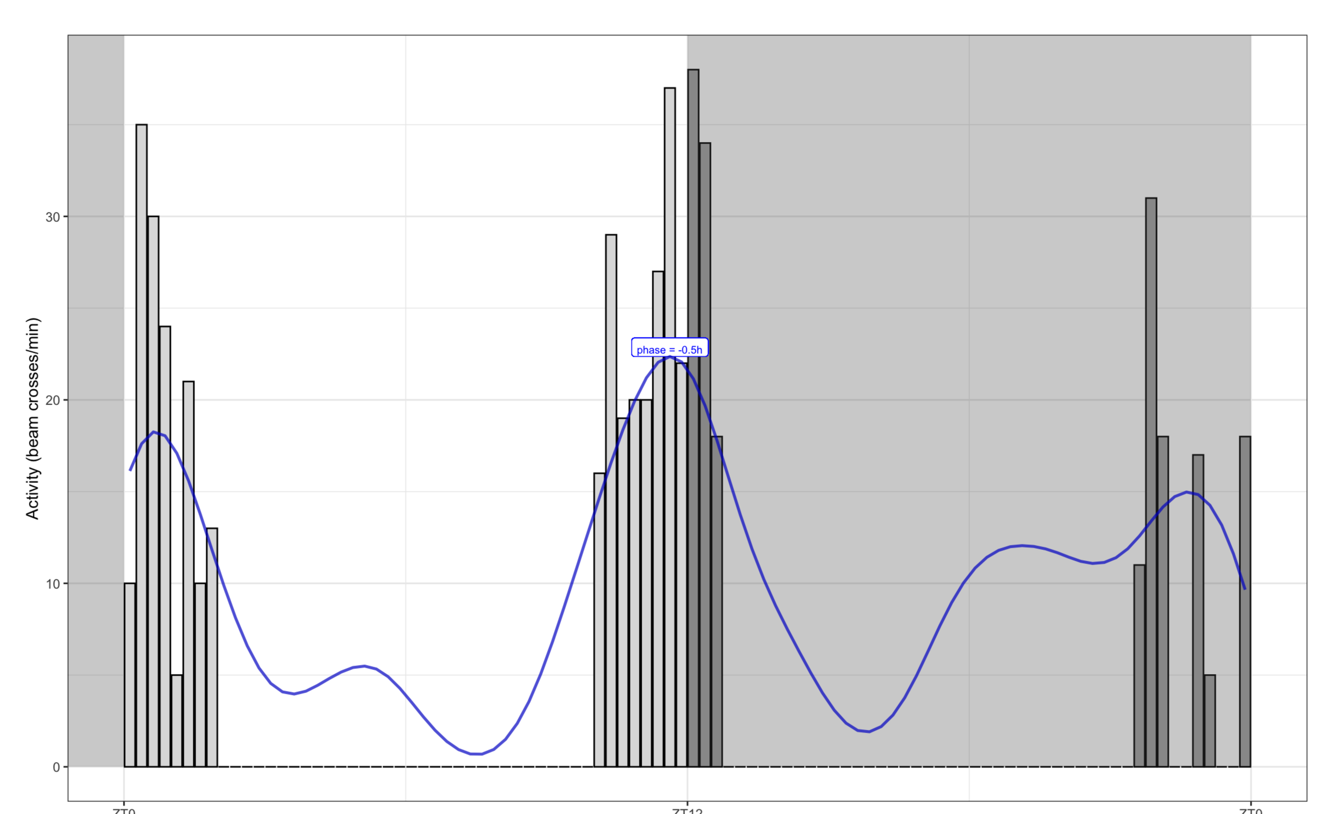

Figure 2: Evening Phase for fly2 on LD5

Trend Lines

Onset and offset of a peak refer to the start and the end of a peak on the activity plot. Trendlines of a mid-day peak can be obtained by fitting a linear model using linear regression. We use the data between onset & peak, and peak & offset to fit the trendlines, as shown in Figure 3.

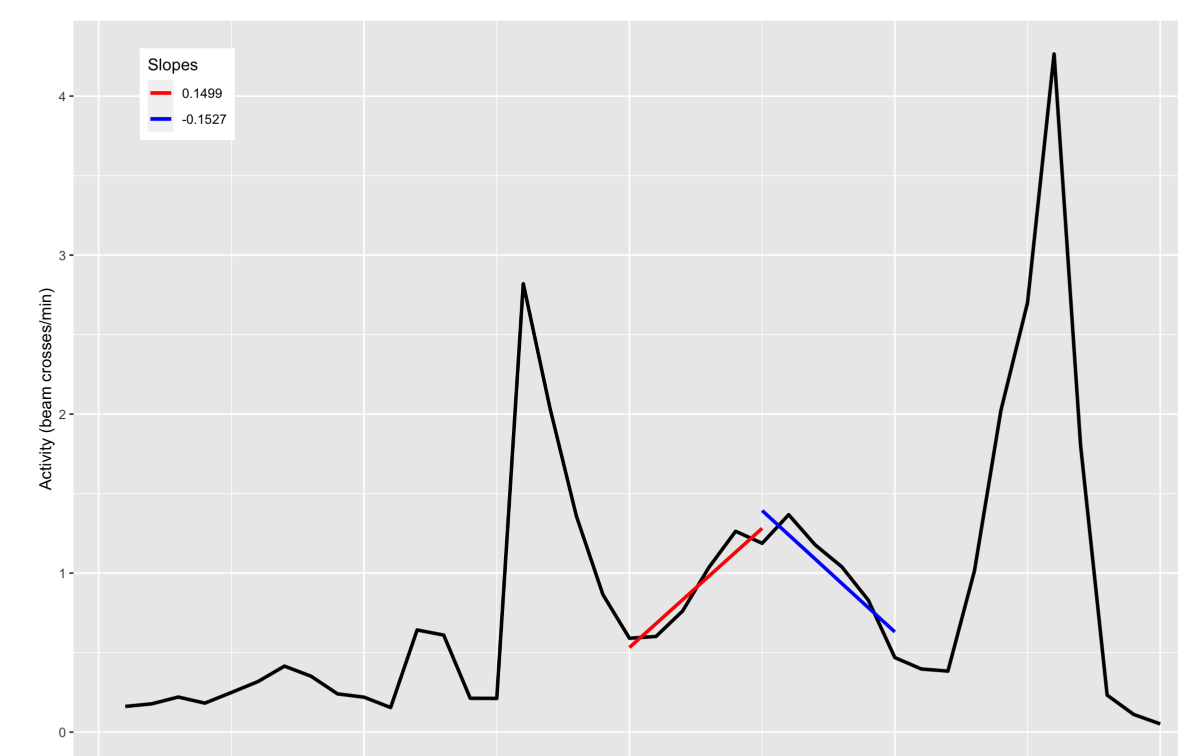

Figure 3: Trendlines for the the mid-day peak.

Shiny Apps

We developed two Shiny Apps for this project. Feel free to explore other components of the project, not covered in the poster, by scanning the **QR Codes** below and accessing the Shiny Apps:

Acknowledgments

We are very grateful to Prof. Maria de la Paz Fernandez, head of the **Neurobiology Lab** at Barnard College of Columbia University, and Prof. Orie Shafer, head of the **Shafer Lab** in the Neuroscience Initiative of the CUNY Advanced Science Research Center. They proposed the projects, provided the data and guided us throughout this interdisciplinary collaboration that we hope to continue in the future for the benefit of all.

References

Fernandez, Maria P., Hannah L. Pettibone, Joseph T. Bogart, Casey J. Roell, Charles E. Davey, Ausra Pranevicius, Khang V. Huynh, Sara M. Lennox, Boyan S. Kostadinov, and Orie T. Shafer. 2020. “Sites of Circadian Clock Neuron Plasticity Mediate Sensory Integration and Entrainment.” *Current Biology* 30 (12): 2225–2237.e5. https://doi.org/10.1016/j.cub.2020.04.025.

Kostadinov, Boyan, Hannah Pettibone, Evardra Bell, Xiaona Zhou, Ausra Pranevicius, Orie Shafer, and Maria Fernandez. 2020. “Open Source Computational Framework in R for Studying Biological Rhythms via Analysis of Behavioral Phase and Circular Statistics.” *STAR Protocols* submitted.