

Graphical Representation of Clustered Functional Actigraphy Data

Abbass Sharif*

Jürgen Symanzik†

Abstract

Human activity and rest patterns could be measured through a device called an actigraph, which collects measurements almost continuously over time. These actigraphy measurements are often treated as functional data. In this article, we first introduce four techniques to visually detect clusters in raw functional actigraphy data for patients with different depression levels. These techniques are: (i) density-based plots, (ii) data enveloping methods, (iii) data summing over time, and (iv) multivariate time series plots. No filtering, smoothing, or any other statistical techniques are required to implement these visualizations. Then, we provide a brief glance at the underlying implementation of these techniques in the R package `ActiVis`. We finish with a brief description of the graphical user interface for this R package.

Key Words: Visualization; Time Series; Data Envelopes; Data Aggregation; Density-based Plots; Data Images.

1. Introduction

Actigraphy is an emerging technology for measuring human activity/rest levels over time. Actigraphy data are collected by an actigraph unit that is a non-invasive watch-like device that consists of an accelerometer. Actigraphy is useful to evaluate sleeping patterns, fatigue, circadian rhythms, and general activity over an extended period of time.

Actigraphy observations are recorded almost continuously over time. Today's actigraphs can measure human activity at different accumulation rates ranging from low (one or more minute intervals) over high (15 seconds intervals) to very high (one second interval). These data could be treated as functional data that can be analyzed with Functional Data Analysis (FDA) techniques (Ramsay and Silverman, 2006). In order to explore and interact with functional actigraphy data, we need some new and easy-to-use visualization techniques and interfaces. In the past, some visualization techniques have been provided by the manufacturers of actigraphs. These techniques have limited features that might not allow the users of the software to view activity levels of one or more subjects from different perspectives. In addition, a user/programmer is not allowed to customize the software provided by the manufacturer to meet his/her needs. In this article, we describe new visualization methods and software based on the object-oriented (OO) programming approach (Lafore, 2002). The resulting R package `ActiVis` will be published as a part of *The Comprehensive R Archive Network-CRAN*, <http://cran.r-project.org>. The main purpose of using the OO paradigm is to make it easier for other programmers to reuse our suggested visualization techniques. For example, this would allow other programmers to create different user interfaces and customize these interfaces for the needs of their users.

*Utah State University, Department of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322-3900, USA. Phone: 435 757 3431, Fax: 435 797 1822, E-mail: abbass.sharif@gmail.com

†Utah State University, Department of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322-3900, USA. Phone: 435 797 0696, Fax: 435 797 1822, E-mail: symanzik@math.usu.edu

The `ActiVis` R package was developed in order to produce reusable statistical tools for visualizing actigraphy data with a user-friendly interface. The American Academy of Sleep Medicine recommended the use of actigraphy as a useful measure for detecting sleep in healthy individuals through assessing specific aspects in insomnia and restless leg syndrome (Ancoli-Israel et al., 2003; Morgenthaler et al., 2007). They also recommended actigraphy as a tool for objectively measuring fatigue. For this purpose — and to overcome the limitations of current visualization tools and software — we developed the software according to the following three specific goals: (i) Development of new multivariate visualization techniques for actigraphy data. (ii) Integration of new and current visualization tools into the `ActiVis` R package using an object-oriented model design. (iii) Development of a user-friendly web interface for actigraphy software.

In Section 2, we provide a brief overview on existing visualization techniques for actigraphy data. The three goals will be discussed in the following three sections. We finish with a brief discussion and outlook in Section 6. More details can be found in Sharif (2012).

2. Current Visualizations of Actigraphy Data

An actigram (sometimes also called actogram) is a commonly used visualization technique for actigraphy data and can be found in numerous actigraphy related publications, e.g., Figures 1 & 2 in Slaven et al. (2009), Figure 2 in Labyak and Bourguignon (2002), and Figure 1 in Sharif et al. (2010). From such a plot, a viewer should be able to detect if there is major minute-to-minute variability during a certain day or significant day-to-day variability among multiple days. Although these graphs allows us to explore the data visually, they only offer limited insight. They are not very powerful tool for studying the variability within a subject and among multiple subjects. For example, if a medical doctor wants to check if a certain treatment or disease is affecting a patient’s activity, then the doctor needs to compare data from before and after the treatment. The actigram won’t allow us to do detailed comparisons. Furthermore, it does not allow us to do group comparisons such as control vs. experimental groups.

Histograms and numerical summary statistics have also been used to explore actigraphy data, e.g., Figure 1(b) in Symanzik and Shannon (2008). These are also not very useful and need to be extended because they describe how patients differ in activity levels, but fail to capture when the patients exhibit different levels of activity levels and patterns.

A variety of new or improved visualization methods for actigraphy data have been suggested in Symanzik and Shannon (2008), such as the raw data plot, smoothed data plot, velocity plot, acceleration plot, cumulative sums plot, and sorted cumulative sums plot. Some of these plots have been previously introduced, such as the cumulative sums plot that resembles the cumulative actigram in Figures 2(B) & 3 in Labyak and Bourguignon (2002). Their implementation in R (R Core Team, 2012) has been presented in Sharif et al. (2010). These visualization techniques for actigraphy data are useful when doing comparisons for a single subject over time (e.g., baseline, during treatment, and after treatment). A detailed description of some of these plots with variation assessment tools can also be found in Ding et al. (2011).

3. Multivariate Visual Data Mining Techniques for Actigraphy Data

As mentioned before, actigraphy is an emerging clinical technology for measuring human sleep, daytime activity, and circadian activity rhythms. Unfortunately, the following three limitations arise while visualizing such types of data, especially when we deal with a large number of subjects and/or many days of data per subject.

- *Information Loss*: Researchers use data smoothing algorithms to fit curves on actigraphy data (Ogbagaber et al., 2012; Wang et al., 2011; Ding et al., 2011; Sharif et al., 2010; Symanzik and Shannon, 2008). Data smoothing is a tool used to reduce the noise and irregularities in order to capture and reveal interesting patterns present in large datasets. However, one of the main drawbacks for data smoothing is losing some insight for the variation in the data. Another critical issue for many of the smoothing algorithms is that they are not robust against outliers, and therefore the smoothed curve is pulled towards those outliers (Rice, 2004).
- *Measurement Bias*: Using actigraphs to measure activity levels of humans might produce substantial measurement error. The actigraphy devices might be biased for many different reasons. In Sherick’s study (Sherick et al., 2010) on the accuracy of the Actical actigraphy devices that are manufactured by the Mini Mitter Company Incorporated (Mini Mitter Company Incorporated, 2005), it was suggested that one of the four sampled Actical devices which were used for measuring activity levels of patients in Sharif et al. (2010) were biased. Another study (Esliger and Tremblay, 2006) on Acticals suggested that even though those devices have small inter- and intra-instrument coefficients of variations, calibration and reliability of devices should not be assumed.
- *Curve Cluttering*: In exploratory data analysis (EDA) (Tukey, 1977), sometimes the researchers might want to compare different sets of groups (males vs. females, different age groups, different races, etc.). Overlaying the smoothed data curves of all groups on the same plot might not produce the desired “clear-cut” grouping of subjects.

In order to overcome the above mentioned limitations we have adopted and enhanced some techniques to help visualize functional datasets, and in particular actigraphy datasets. Four main techniques are introduced: (i) density-based plots such as repeated density strips, (ii) data enveloping methods (*min-max*, 25^{th} – 75^{th} percentile, and 40^{th} – 60^{th} percentile) to summarize common features, (iii) data summing over time (10, 20, 30 and 60 minutes), and (iv) multivariate time series plots such as data images. These techniques are applied to raw data, i.e., unprocessed actigraphy data. No filtering, smoothing, or any other statistical techniques are required at this stage.

3.1 Techniques for Visualizing Functional Data

Consider the scenario in time series data where we have to graph observations recorded every minute over multiple days. This means that we end up having 1440 observations per day (24 hours \times 60 minutes). These data usually have lots of spikes and variations depending on the application. In such scenarios, scientists usually try to visualize the pattern that is present in the data by plotting all observations on one figure where the x-axis (horizontal axis) represents time (in minutes) and

the y-axis (vertical axis) represents the measurement of a particular quantity under investigation. If the scientist wants to compare how those measurements vary among different days, then the traditional way would be to overlay all days on the same plot. This action might produce messy plots with lots of interweaving spikes and lots of data overplotting. Thus, there is no immediate way to distinguish between different days' patterns even if the scientist uses color to differentiate between days. A solution would be to smooth the curves, which has the disadvantage of losing information. In this section, we will present four visualization techniques to explore functional time series data that could be used either alone or combined with some of the other techniques depending on how messy or noisy the data are.

3.1.1 Density-based Plots

The idea of density-based plots is based on Jackson's density strip plots (Jackson, 2008) which display uncertainty with shading. A density strip plot is usually a horizontal rectangle with color shading that ranges from dark to white. It is shaded with darkness proportional to the probability density of the measured quantity at a point, darkest at points of highest probability density, and white at points of zero density. This kind of plots is usually very helpful when comparing distributions arising from parameter estimation.

Density strip plots could be utilized to visualize variability and trends for functional data over time. The main purpose is to make the trends look clearer on the plot without using smoother functions that might lead to loss of information. The proposed density-based plots are simply stacked vertical density strips with equal widths, where each strip shows data for a given period of time (e.g., 1 minute, 10 minutes, 1 hour, etc.). In order to have a proportional shading scheme for all of the strips, the shading level for each strip is multiplied by its density divided by the maximum density over all strips. This technique is similar to Miller and Wegman (1991), where the authors proposed to have a "line density plot".

These density-based plots could be used with either raw data or cumulative sums data. Cumulative sums plots show accumulated activity obtained by adding up activity counts as one moves across the horizontal time axis from midnight (far left) to midnight 24 hours later (far right) (Sharif et al., 2010). They are helpful to show the total activity of a group up to a particular time of the day.

3.1.2 Data Enveloping

Data enveloping is the process of subsetting the dataset into different classes, drawing bands around each class of data observations, and then filling each with a different color. The idea of enveloping data was first introduced by Inselberg et al. (1987) as a tool to reduce noise in Parallel Coordinate Plots (PCPs), and then it was enhanced by Moustafa et al. (2011).

This technique is helpful to clearly see trends followed by each class or visually validate cluster analysis results. These bands often range from the minimum observation to the maximum observation for every minute (*min* – *max* envelopes), but sometimes these extreme observations might be outliers, and thus cause heavy overlapping between the classes. In order to overcome this problem, two things could be done: (1) the bands could be drawn with a narrower range; for example, from the first quartile Q_1 to the third quartile Q_3 of a given class of data, or from the 40th percentile to the 60th percentile, etc., and (2) use alpha blending techniques (Porter

and Duff, 1984) for the colors to create transparency effects in order to be able to see the hidden parts of class bands.

3.1.3 Data Summing

Data summing is the process of combining multiple observations into one single observation. In time series data analysis, we can sum up many observations into 10, 20, 30, or even 60 minute intervals instead of looking at minutely data. This process will help in data reduction, and it automatically reduces the number of spikes and makes the graph look smoother. For example, if we decide to sum up each 10 observations into one observation, we will reduce the daily dataset by 90% (from 1440 to 144). In order to get a better view of data clusters, this technique could be used in combination with data enveloping.

3.1.4 Multivariate Time Series Plots

Multivariate time series plots become handy when we need to compare more than a few time series. In the traditional way, the comparison was done by stacking all time series plots in a fashion where it sometimes become difficult to fit all plots on one page or a computer screen. Peng (2008) visualized environmental data that are collected over time and multiple locations. Instead of producing the traditional time series stacked plots, he came up with a new visualization technique for plotting multivariate time series. This plot is based on the “data images” concept, which was first introduced by Wegman (1990) as colored histograms which are also called “data images” (Minnotte and West, 1998; Morphet and Symanzik, 2010). The idea is simple and similar to density based plots: each time series is split into different categories that are assigned to a color intensity range from low (few observations) to high (many observations). The number of categories varies depending on the nature of the data. This kind of categorization or discretization allows the user to visualize variation in the data.

In addition to the basic data image, multivariate time series plots can be augmented with some summary information at the right and bottom sides of the plot. For example, a box plot could be drawn for each time series at the right hand side of the plot. Also, at the bottom of the plot, we can have some graphs using the previously mentioned techniques which would give summary information about values across all the time series for each specified time point.

3.2 Actigraphy Data

The data used in this article are based on a small sample of 55 patients with insomnia, sleep apnea, or restless leg syndrome, collected at the Washington University in St. Louis Sleep Medicine Center. Two types of data were collected for each patient: actigraphy level data and depression level data. For more information about these data, please refer to Ding et al. (2011). The actigraphy level data were collected via an actigraph device manufactured by the Mini Mitter Company Incorporated Mini Mitter Company Incorporated (2005) that the patient wore on his/her wrist for a period of seven days. Some of the actigraphs collected data every 15 seconds, and others collected data every minute, but for the purpose of this study, we aggregated the 15 seconds level data into one minute level data. The depression level data were collected to investigate patterns in activity levels in different patient subgroups. Each patient filled out the Patient Health Questionnaire (PHQ-9) (Kroenke and

Spitzer, 2002), one of several existing ways to evaluate the level of depression. On the PHQ-9 scale, the higher the depression score, the more depressed the patient is. The following are some descriptive statistics for the collected data in terms of demographics and depression levels:

- Gender: 17 males, 38 females.
- Depression level: 15 patients with no depression (Level 0), 13 with mild depression levels (Level 1), 15 with moderate depression levels (Level 2), 8 with moderately severe depression levels (Level 3), and 4 with severe depression levels (Level 4).

3.3 Visualization Techniques Applied to Actigraphy Data

3.3.1 Density-based Plots

Figure 1 shows the density-based plots for the actigraphy data for groups of patients with different depression levels. These plots look very rugged. It is difficult to compare the activity patterns for these groups of patients. The “disadvantage” of such plots is that we have to separate the groups into different plots. We can see that patients with very high depression levels (Level 4) are more active during the night and have low activity levels early during the morning while the other four groups (Levels 0, 1, 2, and 3) have normal activity patterns — active during the day and passive during the night. This kind of plot does not show us clearly if there is a difference in activity levels for some groups. Thus, to obtain a clearer picture for all groups, we can look at the cumulative sums plots (Figure 2) for the actigraphy data. These plots show accumulated activity obtained by adding up activity counts as one moves across the horizontal time axis from midnight (far left) to midnight 24 hours later (far right). People with higher depression levels accumulate higher activity counts during the night and early morning.

3.3.2 Data Enveloping

Figure 3 shows the actigraphy data of the 55 patients clustered according to the patients’ depression level. Figure 3 (top) shows a 25th – 75th percentile envelope for the raw actigraphy data, while Figure 3 (bottom) shows a narrower 40th – 60th percentile envelope. Figure 3 shows that patients with high depression levels are most active during the night compared to the other groups of patients, and this group is least active during the day. It is not easy to compare the other groups of patients even when we look at a narrower 40th – 60th percentile envelope.

Figure 4 shows the cumulative sums plots for the accumulated sums of actigraphy clustered according to depression levels. Figure 4 (top) uses a 25th – 75th percentile envelope which shows some distinction between the five groups of patients with different depression levels. As anticipated, patients with depression Level 0 have the highest accumulated activity during the whole day, followed by patients at Level 1, then Level 2, then Level 4. It seems that there is a high variability for the Level 3 depression patients. The 40th – 60th percentile envelope (Figure 4 (bottom)) shows that there might be an outlier in the group with depression Level 3 that was affecting the variability of this group. Opposite to what we anticipated, patients that were diagnosed with depression Level 3 have the highest activity. This result might be due to the small sample of just 8 patients with depression Level 3.

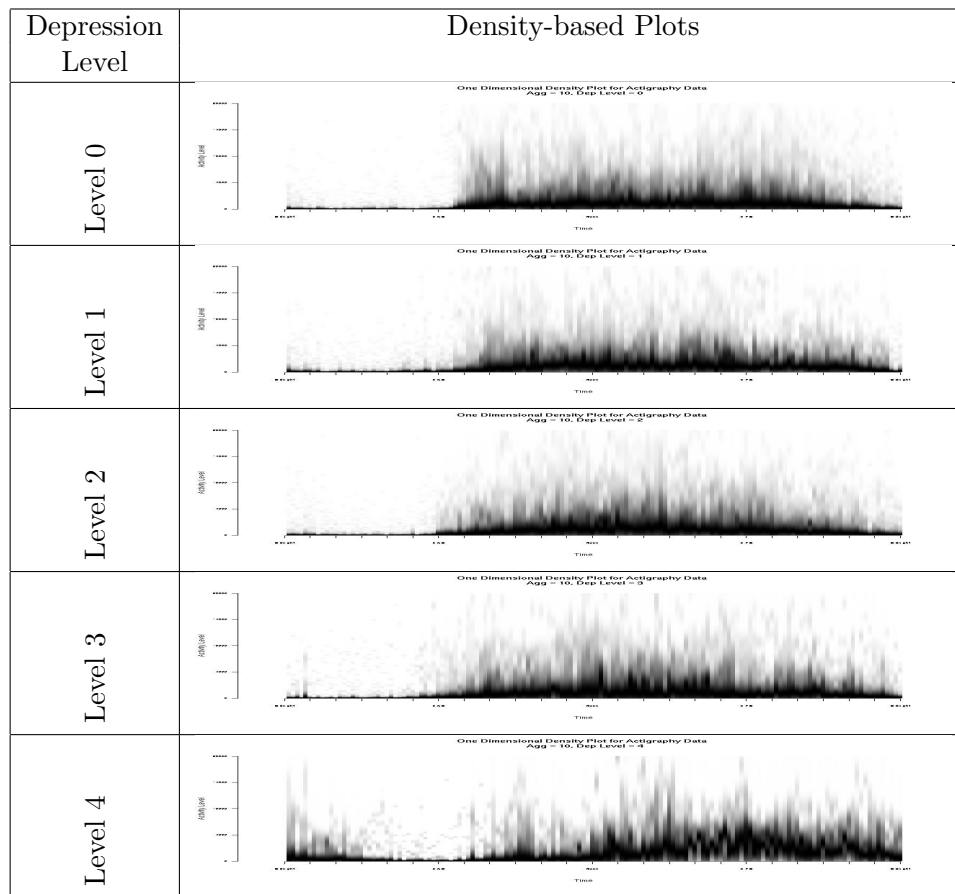


Figure 1: Density-based Plots: Actigraphy data grouped by patients' depression levels.

3.3.3 Data Summing

In Figure 3, the actigraphy plot had lots of spikes and it was hard to see the patterns for each class of patients even after we applied enveloping techniques. To make the plots look smoother and obtain a better separation among different patients' groups, we aggregated the data and summed it over time in addition to using enveloping techniques. Figure 5 (top) shows 25th – 75th percentile envelopes, but with data summing every 20 mins. In other words, the activity levels for each patient are accumulated into one observation every 20 mins. This technique, combined with enveloping, helps to smooth the plot. Figure 5 (bottom) shows the actigraphy data in a 40th – 60th percentile envelope for all patients' groups based on depression levels.

3.3.4 Multivariate Time Series Plots

Figure 6 shows the multivariate time series plot for all patients. There are 55 horizontal time series strips in the image plot (top left), divided into five depression categories that are separated with bold horizontal black lines. For this dataset, it is not clear whether patients with different depression levels have different activity levels due to the high noise. From the box plots on the right hand side of the image plot, we can see that the variation in activity levels for some patients is high. This

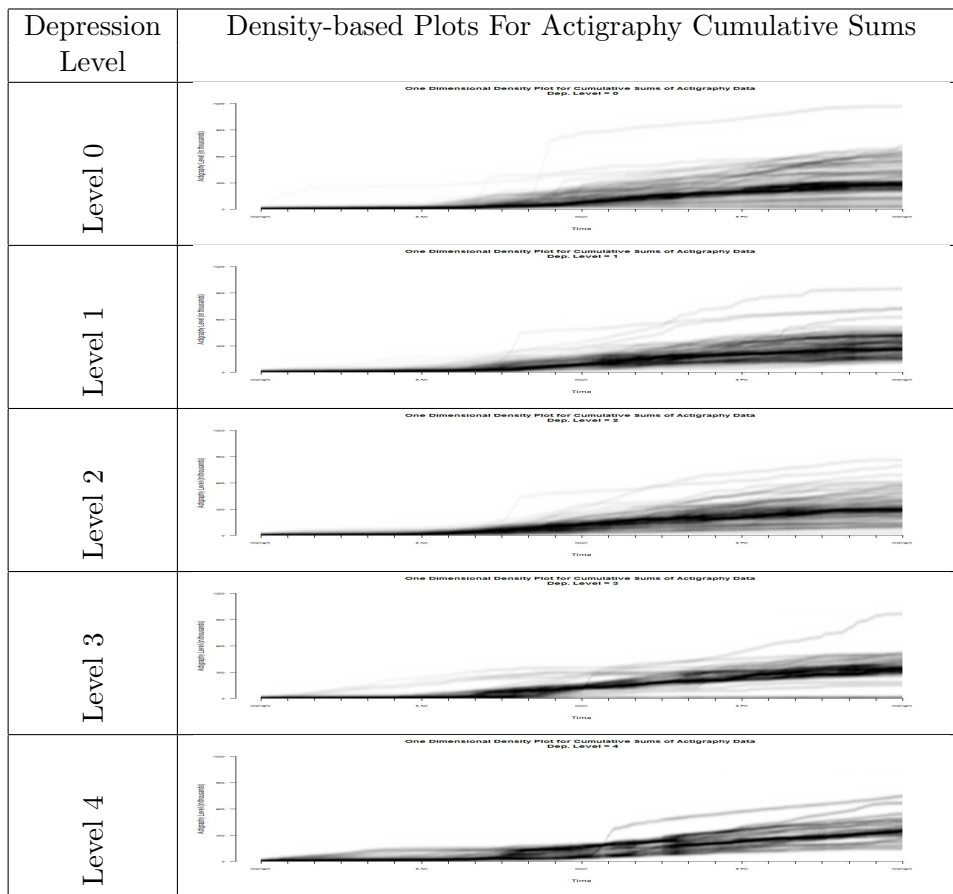


Figure 2: Density-based Plots: Cumulative sum plots for actigraphy data grouped by patients’ depression levels.

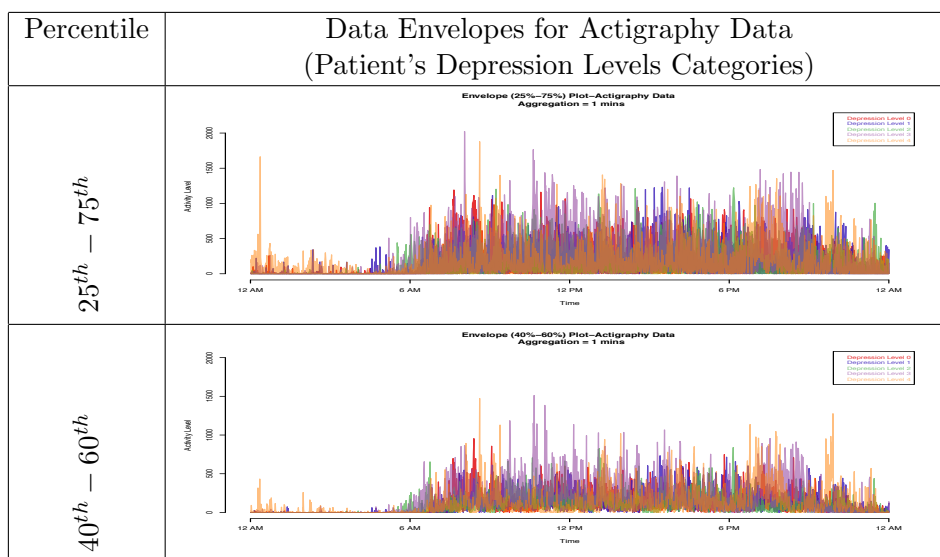


Figure 3: Data Envelopes: 25th – 75th percentile vs. 40th – 60th percentile envelopes for actigraphy data grouped by patients’ depression levels.

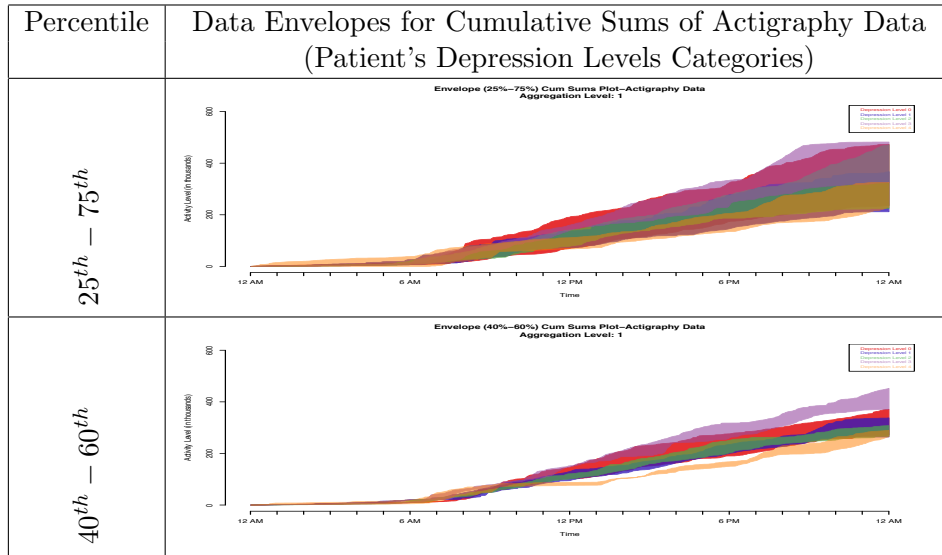


Figure 4: Data Envelopes: 25th – 75th percentile vs. 40th – 60th percentile envelopes for cumulative sums of actigraphy data grouped by patients' depression levels.

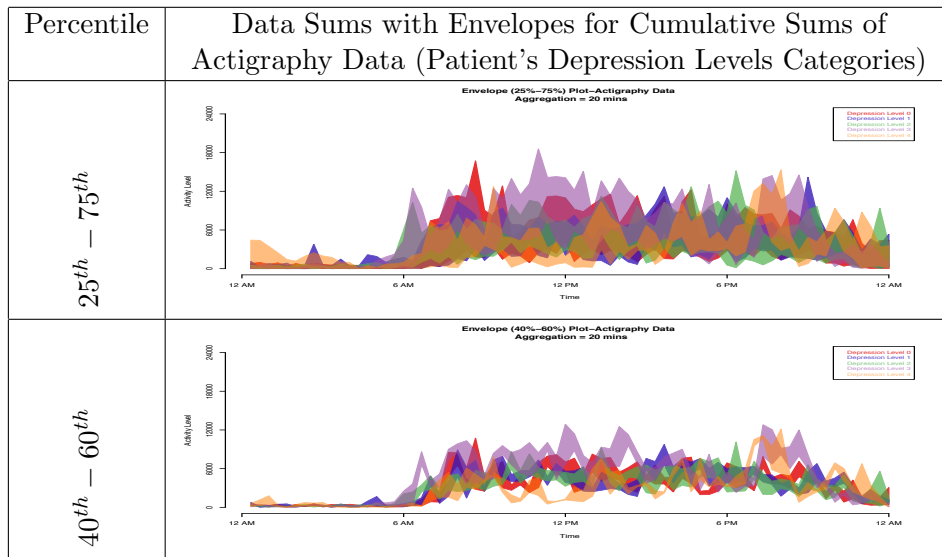


Figure 5: Data Sums with Envelopes: 25th – 75th percentile vs. 40th – 60th percentile envelopes for actigraphy data grouped by patients' depression levels.

might be due to the presence of outliers.

4. ActiVis: An R Package for Visualizing Actigraphy Data

A preliminary object-oriented (OO) software implementation for the visualization of actigraphy data has been described in Sharif et al. (2010). Here, we introduce a full model that also utilizes the OO programming approach (Lafore, 2002). The main purpose for using the OO paradigm is to make it easier for other programmers to reuse our suggested visualization techniques. For example, this would allow other programmers to create different user interfaces and customize these according to the

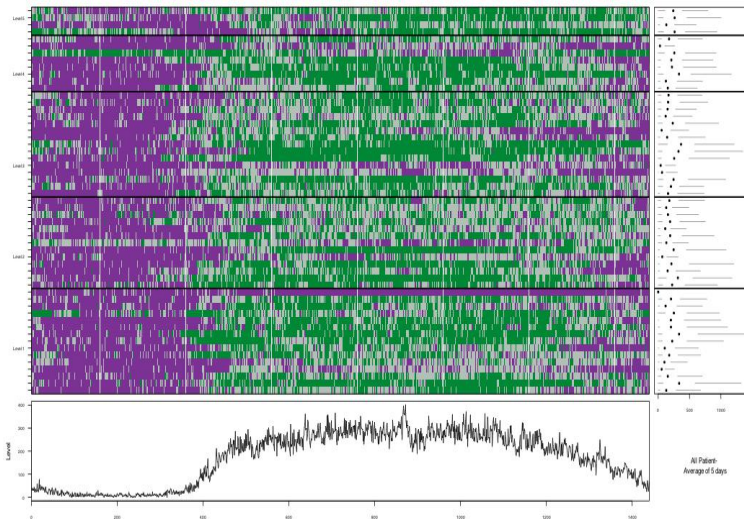


Figure 6: Multivariate Time Series Plot: Actigraphy data with different levels of depression — Level 0 (bottom) to Level 4 (top). Purple represents low and green represents high activity levels. The right hand side plot shows a box plot, and the bottom side shows a plot for the median activity level across all the times series for each time point.

needs of different users.

Object-oriented programming is a way of thinking differently from the widely used procedural programming approach. Thus, instead of implementing data structures and procedures, the programmer now implements *objects* and *methods* used by those objects. In the OO approach, data are combined with *methods* (which are supposed to have access to the data) into one single component called an *object*. This makes any unauthorized access to the data by a different component of the software almost impossible. The OO approach binds data closely to the methods that operate on them and protects the data from any accidental modifications from outside methods. The problem with procedural programming lies in the separation between the data and procedures, because this provides a poor model for real world applications (Lafore, 2002).

The ActiVis R package consists of eleven classes. These are outlined in the class diagram in Figure 7. A class diagram displays the structure of the system by showing its classes, their attributes and methods, and the relationships between the classes.

In order to use the visualizations from the ActiVis package, the user needs to set up a *Graph* object. The following chunk of code initializes a *Graph* object of class *RawDataPlot* for Karli’s object of class *ActData*, and then plots it:

```
#create a new raw data plot object
karli_raw <- RawDataPlot$new()
karli_raw$legendPosition <- "topright"
karli_raw$act_range <- c(0, 4000)
#setup the graph by calling the graphics device
karli_raw$setup()
#show the average of data for the days
karli_raw$avgPlot <- T
#what days to plot?
karli_raw$plotdays <- 2:3
#store Karli’s object in ‘patient’ field
```

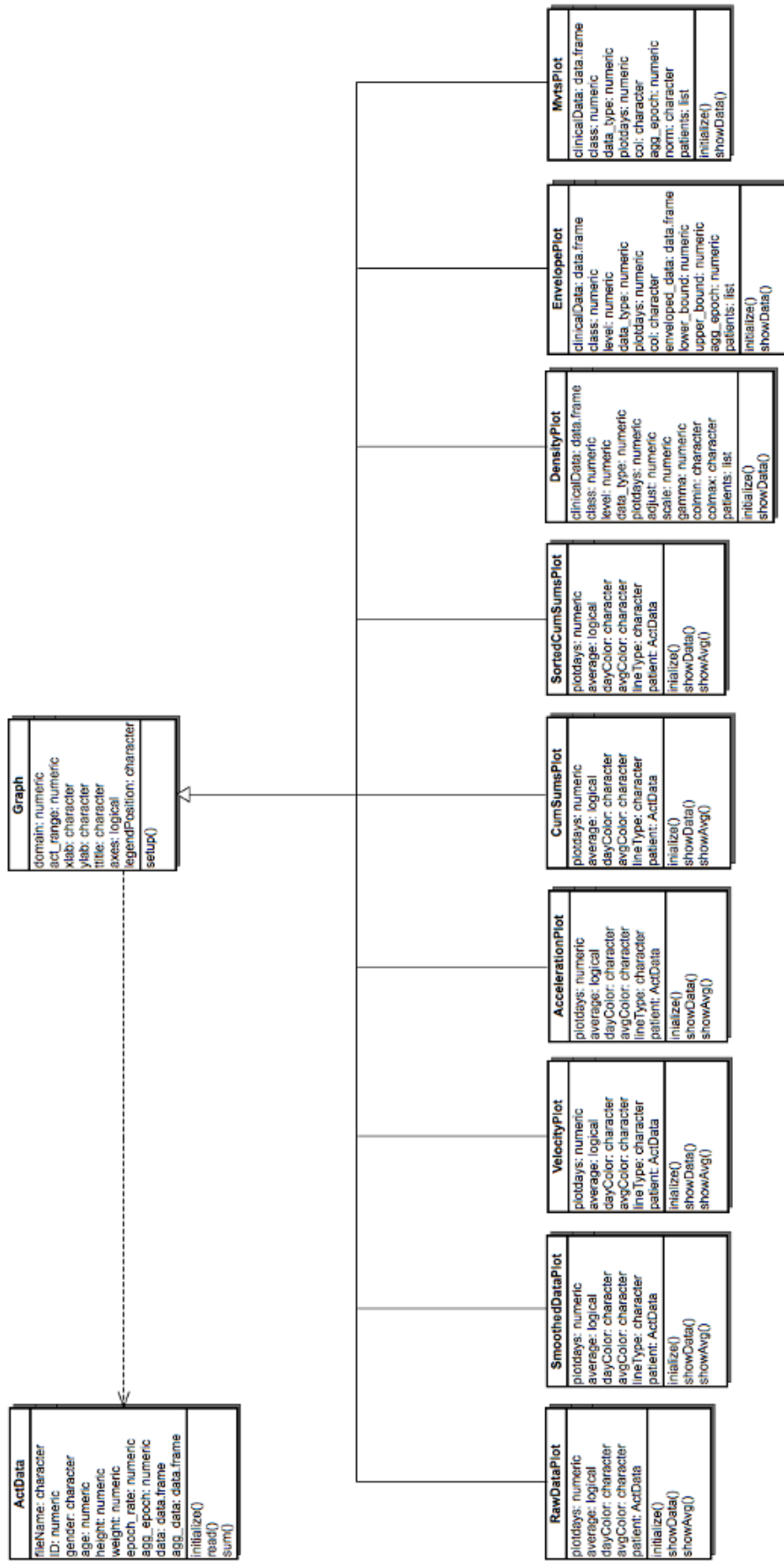


Figure 7: Class diagram for the Act.iVis R package. Shown are eleven different classes and the interactions among them. The dashed arrow represents a dependency relation, while the non-dashed arrows represent inheritance relations. Inheritance indicates that one of the two related classes is considered to be a specialized form of the other, while a dependency relationship indicates that one class depends on another because it uses it at some point of time.

```

karli_raw$showData()
karli_raw$patient <- karli
#show the data on the raw data plot
karli_raw$showData()

```

Instantiating and plotting the other plot objects could be done similarly to the *RawDataPlot* in this example.

5. A Web-based User Interface for ActiVis

Many end users of the ActiVis R package can be expected to have a background in the medical field, sports, or they might be individuals interested in observing their daily activity levels. Those users are unlikely to know how to deal with computer code written in R. Furthermore, R does not have a user-friendly Graphical User Interface (GUI) with menus and buttons. These users most likely want to focus on the results and graphics, and not on running computer code. In this section, we discuss an adaptive web interface that facilitates the use of our ActiVis R package without the need to write or run any R code.

Most of the existing statistical web applications of R are simple interfaces to the R environment. Many of them (e.g., *Rweb* and *webbioc*) allow users to write R scripts and submit them through their interfaces. A few others have more specific purposes with a user-friendly GUI, such as the web interface by Ooms (2010) to the *ggplot2* R package (Wickham, 2009). Our ActiVis web interface was inspired by Oom’s web applications and interfaces to R (Ooms, 2009, 2010). Our GUI consists of two main parts: (1) the left hand side is the user interaction area, and (2) the right hand side is the plot display area.

The user interaction area is designed in a way to make it easy for the user to follow the steps needed for producing actigraphy plots. For this purpose, we used an “accordion” GUI control, which is a stacked list of items that could be expanded either automatically or manually to reveal the content associated with each option. The user interaction area for the actigraphy web interface has three items: File Upload (shown), Graph Type (hidden), and Graph Parameters (hidden). Figure 8 shows a snapshot of the application after the user has clicked on the “Browse” button: A file browser window pops up and prompts the user to select one or many actigraphy data files for uploading. The second file upload field is optional, and it is for uploading the PHQ-9 scores in case the user wants to do cluster visualization. After the selection of files is done, the “Upload” button becomes active, and the user can click on it to submit his/her request.

In order to keep the interface clean and clear, we decided to make it adaptive. In other words, the interface presents the “next” steps based on the user’s input. For example, if the user uploads only one data file, this means that he/she is planning to do visualizations for only one patient. Thus, the plots for one patient visualization will show up as options. On the other hand, if more than one file is uploaded, the system will automatically detect that the user is planning to do cluster analysis, and thus, multivariate visualization options will show up.

The transmission between the three steps of user interaction is smooth and automatic. When the system finishes uploading the files to the server, the “Graph Type” area of the accordion GUI control (Figure 9) slides up in an animated way and hides the “File Upload” area. This ensures that the user knows that the uploading is done, and in addition to that, he/she will be notified how many files have been uploaded at the title of the “File Upload” section. This kind of feedback is important to keep the users updated and active.

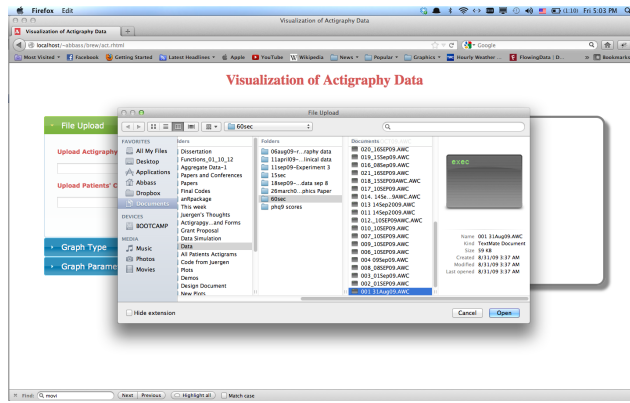


Figure 8: Actigraphy Web Application: Browsing for Data Files.

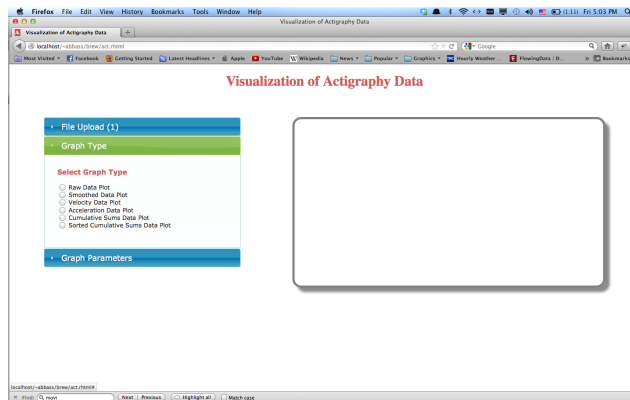


Figure 9: Actigraphy Web Application: Selecting a Graph Type.

In this example, the user has uploaded one file. The visualization options given to the user are for examining a single patient’s behavior and not for cluster analysis. When the user selects a graph type, the “Graph Parameter” area of the accordion GUI control slides up and hides the “Graph Type” area, and the system shows a feedback message to the user about the type of visualization selected.

The third and last step before rendering and showing the graph to the user is to get some graph parameters. Currently, the parameters include the title of the plot and the range for the data. These parameters are optional, and if the user decides not to enter them, then the default parameters are used. When the user has submitted all parameters and is waiting for the graph to show up, a progress bar is shown beside the “Plot” button to provide feedback to the user while the server is doing the computations.

When the server is done with the computations, the progress bar disappears, and the selected graph for the uploaded data appears on the graph display area of the web interface (Figure 10). If the user double-clicks on the graph, a new window will pop up showing only that graph. Users also have the option of downloading the graph to their local hard drive.

Because of the technologies used for this web application, the user does not have to start from the beginning if he/she decides to obtain a different plot for the uploaded data. The user can click on the “Graph Type” accordion GUI control area, and change the choice, and then the system will automatically take him/her through the necessary steps to produce the selected graph. This kind of interaction

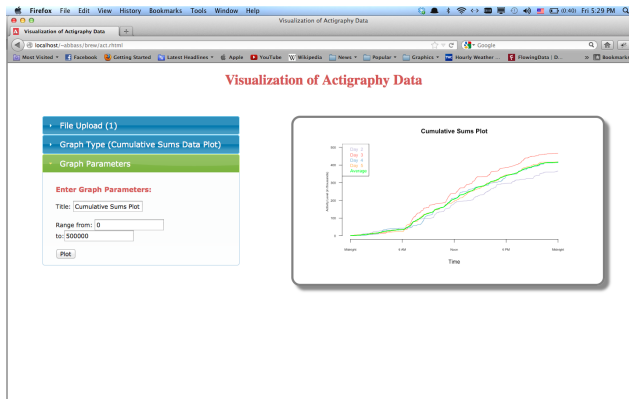


Figure 10: Actigraphy Web Application: Showing the Visualization.

makes the web application dynamic and adaptive, and at the same time efficient and easy to use because the user does not have to reload the whole application to change the graph type or even one parameter in the chosen graph.

6. Summary and Conclusions

In this article, we presented visualization techniques for exploring functional actigraphy data. In particular, we looked into ways to visualize the activity of patients who might be suffering from depression, and noticed how their activity levels varied from one day to another within one patient as well as among different patients during different days. These visualization techniques were implemented in R following the object-oriented paradigm in order to facilitate the process of updating and reusing the developed computer code in the future. We also bundled our techniques into an “open source” R package which will be available on the CRAN website sometime soon. Finally, and to make our system accessible by a wide variety of audiences, we developed a web application with a graphical user interface that deploys all of our techniques.

One of the limitations that we faced in producing our multivariate plots (density, envelope, and multivariate time series plots) is the speed in processing the necessary computations. This speed would be even slower when the computations were done via the web application because it involves an extra step of transferring the data over the network. We have data for 55 patients, and on average each patient has 12 days of data. The data collection was done every minute, so in total our dataset has almost 1 million observations ($55 \text{ patients} \times 12 \text{ days} \times 1440 \text{ minutes}$). The process of data clustering takes even more time when we look at a representative sample of 750 patients as planned for the full study in Ding et al. (2011).

Visualizing functional actigraphy data gives us an insight of how activity levels of patients vary. The natural step that comes after visualization is analyzing these data, and one of our future goals is to quantitatively investigate the source of variation. For this purpose, we already started looking at FDA techniques, and in particular at Functional Principle Component Analysis (FPCA) of actigraphy data (Ding et al., 2011). This work was conducted in collaboration with researchers at the Washington University in St. Louis Sleep Center, who also created an R package called **Actigraphy** (Shannon et al., 2012) for this purpose. The techniques presented in this article could be adapted to visualize any kind of functional data, and not just actigraphy data.

References

- Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W. and Pollak, C. (2003), ‘The Role of Actigraphy in the Study of Sleep and Circadian Rhythms’, *Sleep* **26**(3), 342–392.
- Ding, J., Symanzik, J., Sharif, A., Wang, J., Duntley, S. and Shannon, W. D. (2011), ‘Powerful Actigraphy Data Through Functional Representation’, *Chance* **24**(1), 30–36.
- Esliger, D. W. and Tremblay, M. (2006), ‘Technical Reliability Assessment of Three Accelerometer Models in a Mechanical Setup’, *Medicine & Science in Sports & Exercise* **38**(12), 2173–2181.
- Inselberg, A., Chomut, T. and Reif, M. (1987), ‘Convexity Algorithms in Parallel Coordinate’, *Journal of the ACM* **34**(4), 765–801.
- Jackson, C. (2008), ‘Displaying Uncertainty with Shading’, *The American Statistician* **62**(4), 340–347.
- Kroenke, K. and Spitzer, R. L. (2002), ‘The PHQ–9: A New Depression Diagnostic and Severity Measure’, *Psychiatric Annals* **32**(9), 509–515.
- Labyak, S. E. and Bourguignon, C. (2002), ‘Measurement Issues Related to Actigraphy Use in Older Women’, *Topics in Geriatric Rehabilitation* **18**(1), 68–79.
- Lafore, R. (2002), *Object-Oriented Programming in C++*, Sams Pub., Indianapolis, IN.
- Miller, J. J. and Wegman, E. J. (1991), Construction of Line Densities for Parallel Coordinate Plots, in A. Buja and P. A. Tukey, eds, ‘Computing and Graphics in Statistics’, Springer, New York, NY, pp. 107–123.
- Mini Mitter Company Incorporated (2005), *Actical Physical Activity Monitoring System Instruction Manual*, Bend, OR.
- Minnotte, M. C. and West, R. W. (1998), The Data Image: A Tool for Exploring High Dimensional Data Sets, in ‘1998 Proceedings of the Section on Statistical Graphics’, American Statistical Association, Alexandria, VA, pp. 25–33.
- Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson, A., Coleman, J., Lee-Chiong, T., Pancer, J. and Swick, T. J. (2007), ‘Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007’, *Sleep* **30**(4), 519–529.
- Morphet, W. and Symanzik, J. (2010), ‘The Circular Dataimage, a Graph for High-resolution Circular-spatial Data’, *International Journal of Digital Earth* **3**(1), 47–71.
- Moustafa, R. I., Hadi, A. S. and Symanzik, J. (2011), ‘Multi-Class Data Exploration Using Space Transformed Visualization Plots’, *Journal of Computational and Graphical Statistics* **20**(2), 298–315.
- Ogbagaber, S., Albert, P. S., Lewin, D. and Iannotti, R. J. (2012), ‘Summer Activity Patterns Among Teenage Girls: Harmonic Shape Invariant Modeling to Estimate Circadian Cycles’, *Journal of Circadian Rhythms* **10**(2).
- Ooms, J. (2009), rweb.stat.ucla.edu/irttool/: A Web Interface for the R Package IRT.
URL: <http://rweb.stat.ucla.edu/irttool/>
- Ooms, J. (2010), [yeroon.net/ggplot2](http://www.yeroon.net/ggplot2/): A Web Interface for the R Package ggplot2. Version 0.2.
URL: http://www.yeroon.net/ggplot2
- Peng, R. (2008), ‘A Method for Visualizing Multivariate Time Series Data’, *Journal of Statistical Software* **25**(1), 1–17.
- Porter, T. and Duff, T. (1984), ‘Compositing Digital Images’, *SIGGRAPH Computer Graphics* **18**(3), 253–259.
- R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org/>
- Ramsay, J. O. and Silverman, B. W. (2006), *Functional Data Analysis (2nd Edition)*, Springer, New York, NY.
- Rice, J. A. (2004), ‘Functional and Longitudinal Data Analysis: Perspectives on Smoothing’, *Statistica Sinica* **14**(3), 631–647.
- Shannon, W., Li, T., Xian, H., Wang, J., Deych, E. and Gonzalez, C. (2012), *Actigraphy: Actigraphy Data Analysis*. R package version 1.0.
URL: <http://CRAN.R-project.org/package=Actigraphy>
- Sharif, A. (2012), Visual Data Mining Techniques for Functional Actigraphy Data: An Object-Oriented Approach in R, PhD thesis, Utah State University, Logan, Utah, USA.
- Sharif, A., Symanzik, J. and Shannon, W. D. (2010), ‘An Object-Oriented Approach in R for the Visualization of Functional Actigraphy Data’, *Computing Science and Statistics* **41**. Forthcoming.
- Sherick, P., Symanzik, J. and Shannon, W. (2010), Accuracy, Bias, and Notable Observation for Experimental Actigraphy Data, in ‘2011 JSM Proceedings’, American Statistical Association, Alexandria, VA. (CD).
- Slaven, J. E., Andrew, M. E., Mnatsakanova, A., Violanti, J. M., Burchfiel, C. M. and Vila, B. J. (2009), ‘Statistical Modeling of Sleep’, *Chance* **22**(1), 16–21.
- Symanzik, J. and Shannon, W. (2008), Exploratory Graphics for Functional Actigraphy Data, in ‘2008 JSM Proceedings’, American Statistical Association, Alexandria, VA. (CD).
- Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison Wesley, Reading, MA.
- Wang, J., Xian, H., Licis, A., Deych, E., Ding, J., McLeland, C., Li, T., Duntley, S. and Shannon, W. (2011), ‘Measuring the Impact of Apnea and Obesity on Circadian Activity Patterns Using Functional Linear Modeling of Actigraphy Data’, *Journal of Circadian Rhythms* **9**(11).
- Wegman, E. J. (1990), ‘Hyperdimensional Data Analysis Using Parallel Coordinates’, *Journal of the American Statistical Association* **85**, 664–675.
- Wickham, H. (2009), *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York, NY.