A peer-reviewed version of this preprint was published in PeerJ on 3 February 2020.

View the peer-reviewed version (peerj.com/articles/8341), which is the preferred citable publication unless you specifically need to cite this preprint.

Garnier EM, Fouret N, Descoins M. 2020. ViSiElse: an innovative R-package to visualize raw behavioral data over time. PeerJ 8:e8341 https://doi.org/10.7717/peerj.8341

PeerJ Preprints NOT PEER-REVIEWED

ViSiElse: An innovative R-package to visualize raw behavioral data over time

Elodie M Garnier Corresp., [1] , **Nastasia Fouret** [1] , **Médéric Descoins** Corresp. [1, 2]

[1] Centre d'Études Périnatales de l'Océan Indien (CEPOI, EA 7388), Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion

[2] Centre de Simulation en Santé de l'Océan Indien, Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion

Corresponding Authors: Elodie M Garnier, Médéric Descoins
Email address: e.garnier30@gmail.com, mederic.descoins@chu-reunion.fr

The scientific community encourages the use of raw data graphs to improve the reliability and transparency of the results presented in articles. However, the current methods used to visualize raw data are limited to one or two numerical variables per graph and/or small sample sizes. In the behavioral sciences, numerous variables must be plotted together in order to gain insight into the behavior in question. In this paper, we present ViSiElse, an R-package offering a new approach in the visualization of raw data. ViSiElse was developed with the open-source software R to visualize behavioral observations over time based on raw time data extracted from visually recorded sessions of experimental observations. ViSiElse gives a global overview of a process by creating a visualization of the timestamps for multiple actions and all participants into a single graph; individual or group behavior can then be easily assessed. Additional features allow users to further inspect their data by including summary statistics and time constraints.

PeerJ Preprints NOT PEER-REVIEWED

ViSiElse: An innovative R-package to visualize raw behavioral data over time

Elodie M. Garnier[1], Nastasia Fouret[1] and Médéric Descoins[1, 2]

[1] Centre d'Études Périnatales de l'Océan Indien (CEPOI, EA 7388), Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion.
[2] Centre de Simulation en Santé de l'Océan Indien, Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion.

Corresponding Authors:
Elodie Garnier[1]
Centre d'Étude Périnatales de l'Océan Indien (CEPOI, EA 7388), Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion.
Email address: e.garnier30@gmail.com

Médéric Descoins[1, 2]
Centre d'Étude Périnatales de l'Océan Indien (CEPOI, EA 7388), Centre Hospitalier Universitaire de La Réunion, Saint-Pierre, La Réunion.
Email address: mederic.descoins@chu-reunion.fr

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019

Abstract

The scientific community encourages the use of raw data graphs to improve the reliability and transparency of the results presented in articles. However, the current methods used to visualize raw data are limited to one or two numerical variables per graph and/or small sample sizes. In the behavioral sciences, numerous variables must be plotted together in order to gain insight into the behavior in question. In this paper, we present ViSiElse, an R-package offering a new approach in the visualization of raw data. ViSiElse was developed with the open-source software R to visualize behavioral observations over time based on raw time data extracted from visually recorded sessions of experimental observations. ViSiElse gives a global overview of a process by creating a visualization of the timestamps for multiple actions and all participants into a single graph; individual or group behavior can then be easily assessed. Additional features allow users to further inspect their data by including summary statistics and time constraints.

Introduction

Time data are temporal observations acquired from different sources like video-recorded experiments, sensors, web navigation, or direct measurements; this type of data is used in many research fields including economics, biology, medicine, and the social sciences. In this paper, we use the term “raw time data” to refer to any timestamp directly extracted from its sources without any transformation. On the contrary, the term “non-raw time data” refers to processed or summarized time data.

Behavioral science focuses on the development of behavioral knowledge, represented by a series of actions, through experimental observations. Analyzing action timestamps allows researchers to determine whether the observed behavior is appropriate. A correct behavior is determined by the series of actions that lead to the achievement of a goal. Large amounts of raw time data must be sorted through to capture the correct behavior. However, this data is stored in large tables, which makes their direct interpretation impossible. Graphical representations make the search for data effortless, instantaneous, and easy to understand. The ideal visualization of behavioral data is a graphic tool that plots the raw time data for each action and for all participants simultaneously. There are few tools that can visualize raw data with several variables on the same graph.

The scientific community values the presentation of raw data in professional publications as open science becomes more popular and investigators are encouraged to become more transparent (Fosang & Colbran, 2015; Prager et al., 2018; Rousselet, Foxe, & Bolam, 2016). Studies have demonstrated that some methods that graphically summarize data may be misleading or suggest conclusions that are contrary to the actual distribution of data (Weissgerber, Milic, Winham, & Garovic, 2015). Plotting raw data increases clarity and makes results more understandable and reliable. It is important to choose a relevant plot to correspond with the data and to favor graphics that present all of the data and their structures. Scatter plots, violin plots, beeswarms, or pirate plots present the full range of data and are better choices than those that only present a summary, such as bar- or line plots (Hertel, 2018; Larson-Hall, 2017; Pastore, Lionetti, & Altoè, 2017). Allen et al. (2018) introduced raincloud plots, which are an easy-to-use multi-platform tool that combines visual representations and provides a complete overview of the data with robust, transparent plot. Other researchers turned to interactive visualizations (Ellis & Merdian, 2015; Goedhart, 2018; Weissgerber et al., 2017), allowing readers to explore the dataset with customizable graphs, and measures of central tendency and errors. These solutions successfully answer the need for reliability and transparency in publications. However, the visualization of raw data is often limited to one or two numerical variables per graph and/or small sample sizes. There are no known models that include an innovative raw time data visualization tool displaying an entire process for large samples of participants in a single graph.

ViSiElse is a graphical tool developed to fill the need for such a model and can provide a visualization and global insight of individual and/or group actions over time. ViSiElse was developed with the statistical programming language R (R Core Team, 2018) and is available on the Comprehensive R Archive Network (CRAN) web site: https://CRAN.R-project.org/package=ViSiElse. This program allows for the visualization of raw time data extracted from any experimental observation. The package includes options for additional graphical information of the data tendency (mean, standard deviation, quartiles, or statistical tests) and there are time constraints for each action that check the accuracy of the realized actions. ViSiElse offers a new solution for data reliability and transparency with the visualization of a complete raw dataset in a single graph.

We will provide a step-by-step presentation of the main features of ViSiElse, describing how to set up the data to create a ViSiElse plot, how to customize this plot to get a clear view of the individuals' actions, and how to add time limits or statistical measurements. An example using the actions performed on a typical day will provide a simulated dataset to illustrate the use of this program. ViSiElse is compared with other raw data visualization tools and the various applications and the range of possible uses of ViSiElse are discussed.

The online supplementary material contains the R code needed to reproduce the presented results. Two vignettes (R online documentation) are available for this package; the first describes the ViSiElse process step-by-step using the example of how to make coffee (https://cran.r-project.org/web/packages/ViSiElse/vignettes/ViSiElSe_Step_by_Step.html) and the second follows the example and R script introduced in this paper (https://cran.r-project.org/web/packages/ViSiElse/vignettes/ViSiElSe_Paper_Walkthrough.html).

Methods

The first step in the process of ViSiElse is to define the manner of the observed behavior, build the dataset, and then create an R object, known as the ViSibook.

Creation of the dataset and ViSibook objects

In order to build the dataset correctly, researchers have to a priori translate a behavior or procedure into a linear process of actions. This process has been described by Almeida & Azkune (2018) and demonstrates how behavior is deconstructed into activities and then activities into actions. A list of actions, their type, and their order must be defined in the ViSiElse program. The fundamental question to answer is: “What are the elementary actions comprising the behavior?”. The elementary action is defined as an action that cannot be divided into shorter actions with regard to the time scale. For example, a typical set of daily tasks can be divided into the following elementary actions: sleep, wake up, shower, eat breakfast, drink the first coffee of the day, start and stop working, lunch break, pick up the kids, cook and eat dinner, and then go to sleep. Van Kasteren et al. (2008) used a similar deconstruction of daily activities but their studies only focused on actions that were completed at home.

Once the list of elementary actions is established the actions should be classified as being punctual or long. A punctual action is an action with no duration or a duration that is not long enough to be measured regarding the time scale of the studied behavior. A long action is an action having a duration defined by two punctual actions, one of which occurs at its beginning and one at its ending. For example, the action “sleep” is long while the action “wake up” is punctual. We can also add new actions like the long action “working” which is delimited by the two punctual actions “start working” and “stop working”.

Finally, to have a linear process, actions should be chronologically sorted and numbered; however, if two actions are happening simultaneously they should be ranked randomly. It is not relevant to assign a rank to an action if that action is defined as a punctual action meant only to indicate the start or stop of a long action. However, actions that are not ranked will not be plotted.

In our example, the finalized list of actions with their rankings and classifications is:

1. Sleeping—*long*
2. Wake up—*punctual*
3. Take a shower—*punctual*
4. Eat breakfast—*punctual*
5. Start working—*punctual*

6. Working—*long*
7. Stop working—*punctual*
8. Lunch break—*long*
9. Pick up the kids—*punctual*
10. Cook and eat dinner—*long*
11. Go to sleep—*punctual*
12. First coffee—*punctual*

From raw data to datasets. Raw data encompasses the elapsed time from the study's starting point to the completion of each punctual action for all participants. In our example of a typical day, the starting point is midnight and each value is the time elapsed between midnight and the completion of each action (in min). The dataset is the organized raw data in which the first column identifies the individuals and the following columns categorize the punctual actions including those that delimit the long actions. An example of the appropriate data structure is given in Table 1. This dataset is the R-simulated dataset of the typical day example in the ViSiElse package (can be loaded by the command line "data(typDay)").

Building the ViSibook. While the dataset contains the raw time data of the studied behavior, the ViSibook provides its structure in a table consisting of the characteristics of every action. The minimum structure for a ViSibook requires that each action be named, labeled, defined as being punctual or long, ordered in the process, and long actions must be associated with the name of the two punctual actions delimiting its beginning and its ending. The ViSibook can also include time constraints (green and/or black zones) which allows users to check the time accuracy of the realized actions. To create a ViSibook, users can define or import a table, although the users must take care to import the names and the order of the ViSibook's columns (see ViSiElse CRAN documentation https://cran.csiro.au/web/packages/ViSiElse/ViSiElse.pdf). The ViSibook is an optional parameter of the main function, "visielse", which generates a ViSiElse graph of the studied behavior, according to the dataset and the ViSibook. When ViSibook is not specified, "visielse" will compute a default ViSibook from the dataset; when this occurs, the process order is determined by the names of the dataset columns and all actions are automatically determined to be punctual actions. The ViSibook can be extracted from an execution of the "visielse" function at any time and users can modify any information saved in the ViSiBook to adjust the plotted behavior. For example, on the ViSiElse graph, the names of the actions on the y-axis are the labels from the ViSibook, which by default, are the variable names defined by the dataset column names. Variable names are typically brief and lack clarity to describe an action, therefore a more explicit description is preferred on the displayed labels. Users are able to change the action labels in the ViSibook to improve the clarity of the graph.

Visualization of raw time data with ViSiElse

ViSiElse gives an overview for the timestamp distribution of sequential actions for large samples of participants in a single-page graph. This innovative visualization facilitates the comprehension of behaviors based on the profile of the data distribution. ViSiElse also improves the ease of identification for outliers or abnormal behaviors that do not comply with practice recommendations.

Creating the first plot. Running the "visielse" function with a dataset and an optional ViSibook as arguments will create and display the ViSiElse graph. Figure 1 shows a simulated typical day dataset for one hundred participants. Actions on the graph are organized on the y-axis and their executions are distributed along the time axis (x-axis). A rectangle indicates punctual actions accomplished by at least one individual in the specified interval of time. The length of the time interval is set by the breaks on the time axis; the breaks in Figure 1 are set every 30 min from midnight to midnight. The intensity of the color in the rectangles is proportional to the number of individuals who realized the action during the time interval. Long actions are defined by lines with length that are proportional to the duration of the action completed by an individual; lines are chronologically sorted by the action starting time.

To access and adjust the formatting options for the graph, the ViSiElse object should be plotted using the R basic plot function. There are many options for formatting, including changing the size and color of a label, adding a title, or modifying the time interval size and unit, which is set to 10 seconds by default. The users can modify the size of the time interval using the plot function with the scal.unit.tps parameter. However, this will only change the breaks in the time axis and not the size of the time interval that is used to determine the intensity of the color representing the punctual actions, which is calculated by the pixel parameter. Every formatting option is accessible through the plot function while the package features (group comparison, time constrains, statistics) must be defined through the visielse function.

Adjusting the time interval with the pixel parameter. The pixel parameter represents the time precision for punctual actions, which is defined as the time limit for which a subject is moved from one time interval to another. The default pixel size is set to 20 seconds. This value can be adapted to match the time variation of the observed data, with a minimum value of 1 pixel. If users run the plot function with a ViSiElse object, they should verify that the scal.unit.tps parameter in the plot function is the same or smaller than the pixel parameter defined in the ViSiElse object. The plot function changes the formatting option of the ViSiElse graph, so the two parameters should be coherent.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019

Data are aggregated into time intervals. If the pixel parameter is too small then the plotted information will not accumulate enough to allow for interpretation. For example, in Fig 2A, the pixel parameter is set to 10 min, which is too precise to analyze the behavior of activities scaled to fit in a day. Conversely, if the pixel parameter is too large, the plotted information is too crowded to allow for interpretation. In Fig 2B, most of the participants are in the same time interval as the pixel parameter, which is set to 120 min. In this case, we cannot differentiate between participants and therefore we cannot analyze the variation of behavior between them. The pixel parameter must be chosen and tested carefully.

Analysis of raw time data with ViSiElse

ViSiElse offers many features with which to analyze raw behavioral time data. Users may define groups, time constraints, or statistical measurements to complete their graph. ViSiElse assists in the inspection and interpretation of raw time data in a single graph.

Compare group behavior. ViSiElse differentiates between two subsets of participants using color distinctions. The ability to distinguish between experimental groups of participants helps to identify different patterns of behavior. In the example of the typical day dataset, two groups were created: people who employ a babysitter (in blue) and people who do not (in pink). To display groups within ViSiElse, users simply specify the group and the method arguments in the visielse function. The first argument is a vector containing the group distribution for each individual. The second argument is the name of the chosen visualization method. ViSiElse provides three methods with which to plot groups:

- The cut method where each group is represented one under the other in different colors (see Fig 3A). This representation can be used to compare groups as group data are completely graphically dissociated.

- The join method where groups are spatially mixed but are differentiated by distinct colors (see Fig 3B). With this method, users can analyze the group distribution among the data.
- The within method where all data are plotted together in blue and one of the groups is plotted again in pink (see Fig 3.C). This visualization allows users to examine a specific group's behavior against the global population. As the ViSiElse package only allows two colors of distinction, this method is the most suitable option for data containing more than two groups.

Set time constraints. Behavior may be constrained by external guidelines where actions must respect an order and a timing. When this occurs, punctual actions should be placed in a specific period or not be executed before or after a specific time point. Long actions should not exceed a specified duration or continue after a specific time point. ViSiElse uses green and black zones to help visualize these time boundaries. Green zones represent time obligations within which actions should be accomplished. Black zones set time constraints after which actions should not occur. The visual time parameters allows the user to see whether or not the behavior is completed within the appropriate time zone. For each punctual action, users can define one green zone and two black zones (to surround the expected execution times). To create those time zones, users define their delimitations at two-time points, one for the beginning and one for the ending of each zone. The time points of the green and black zones must be defined in the ViSibook object as columns and they are automatically plotted when the visielse function is run. ViSiElse also allows for the repetition of green zones when a punctual action can be achieved in different time zones. For this option, users define the time point of the first green zone in the ViSibook and the time interval between each green zone. For long actions, ViSiElse only offers black zone parameters, which can be restricted by a deadline not to cross or duration not to exceed, indicating the duration of the action. Users must define the time points and the appropriate restriction method in the ViSibook in order to define the time constraints of long actions.

In a typical day, actions are controlled by external rules. For example, the working hours are defined. In our example, people should be at work and start working before 10 a.m., will have a 30-minute break for lunch, and they cannot leave work before 4 p.m. Schools often end at 4 p.m. and close at 5 p.m., leaving a one-hour interval for the child pick-up. Therefore, time constraints are placed on multiple actions to assess if they are completed within the appropriate time zones (Fig. 1). The punctual actions to indicate "start working" and "stop working" each has a black zone for an unacceptable arrival (after 10 a.m.) and departure time (before 4 p.m.), respectively. The punctual action to pick up the kids had one green zone for the acceptable period (from 4 to 5 p.m.) and two black zones for the unacceptable time (outside the one-hour interval). The long action, "lunch break" has a 30 min-duration limitation displayed by a darker blue color.

Analysis with summary statistics. Summary statistics may be added to complete the ViSiElse graph and analyze the tendency of the behavioral data. Users can choose between plotting the mean and standard deviation (Fig. 3) or the median with the first and third quartiles (Fig. 1). ViSiElse will compute a statistical test to compare the time data between the two groups when the summary statistics are defined and the data contains groups (Fig. 3A-3B). ViSiElse runs a Wilcoxon test if the informer parameter is set to mean and standard deviation. However, ViSiElse will run a Mood's two-sample test if the informer parameter is set to median and quartiles. An asterisk appears on the right side of the graph if the statistical test is significant with a 0.01 alpha risk, which is the default value; this value can also be manually set. For example, in

Fig 3A-3B, the significance was set to 0.05, resulting in a significant test for all actions except for the punctual action. ViSiElse performs statistical tests as an indication of the statistical difference between groups. However, ViSiElse is not a reporting tool and only provides the statistical significance of the group comparison. ViSiElse should be supplemented with additional analytical tools and other tests should be run separately in order to get complete results and test details.

Limitation of usual raw data visualization tools

There are many graphical tools available with which to visualize data; three of those methods were selected for comparison and their characteristics are summarized in Table 2.

Scatter plots are commonly used to visualize raw data and are preferred for their ease-of-use with a small number of variables. However, for a highly dimensional dataset, users need to display all variables one by one. For example, Fig 4A shows the 12 graphs required to see the dataset for the punctual actions of the typical day. It is difficult to interpret the scatter plots as there is no global overview of the process and the order of the actions is unclear. To analyze behavioral data, all actions should be plotted together.

One way to plot the data together is to combine all of the scatter plots into a single graph using the violin plot. For example, Fig 4B displays the same 12 punctual actions on both violin and scatter plots. The dots indicate the raw data and the data distribution is illustrated by the violin shape. The violin plots are usually presented vertically, however, we reversed the x and y-axis to keep the time axis horizontal. This visualization provides a global overview of the process. Users can add boxplots to get the data tendency and can display as many groups as are required. The combination of violin and scatter plots is useful for medium-size datasets.

However, an increase in number of variables would interfere with the interpretation of the data as the dots would be too clustered.

Heatmaps are efficient tools for large and highly dimensional datasets. Fig 4C shows the heatmap of the dataset from our example of the typical day. Heatmaps use a gradient of color to indicate the data, like ViSiElse graphs, and can therefore display an unlimited number of participants and a large number of actions. This visualization method allows users to see the global process, the order of the actions over time, and the raw data and data distribution. However, heatmaps do not provide summary statistics or distinguish between groups. The major drawback of using heatmaps, violin plots, and scatter plots is that they only permit punctual actions, meaning that long actions can only be displayed by their start and end times. This is a major limitation when the duration of action matters. Indeed, when punctual actions are plotted, individuals are pulled together so we cannot link a start time to its end time.

Examples of applications

Healthcare procedures

ViSiElse was originally developed to visualize behavioral data extracted from video recorded sessions of simulated healthcare procedures. Medical procedures are frequently taught via high-fidelity simulations to avoid errors and reduce risks to patients that may result from the learning process (Brewin et al., 2015; Kalaniti & Campbell, 2015; Ziv, Wolpe, Small, & Glick, 2006). For example, midwife students are trained in the neonatal resuscitation procedure, including endotracheal intubation (EI). EI is the process of inserting a tube through the mouth and into the airway in order to restore the airway patency of the newborn. EI is a lifesaving procedure and should be readily available to all patients whose ventilation is compromised. EI consists of six punctual actions completed by two long actions:

1- Decision to intubate—*punctual*
2- Stop mask ventilation—*punctual*
3- Insert the laryngoscope blade in the patient's mouth—*punctual*
4- Insert the endotracheal tube—*punctual*
5- Remove the laryngoscope blade out of the patient's mouth—*punctual*
6- Restart to ventilate the patient through the tube—*punctual*
7- Duration of the laryngoscope use—*long*
8- Total duration of the intubation process—*long*

The execution time of each action is extracted from the videotapes of the simulated sessions. EI, like most medical procedures, follows guidelines set by local or international committees, in this case the International Liaison Committee on Resuscitation (ILCOR) (Wyckoff et al., 2015). ViSiElse provides a graphical overview of the EI process and the verification of the adequacy to the recommendations. For example, Fig. 5 shows the EI process performed by 37 midwives students. The dataset is a subset of the data collected from the SIMULRUN 1 project (POE FEDER number RE0001879) that investigated the neonatal resuscitation training of midwives via high-fidelity simulation. All participants gave written informed consent to participate in the study. The study was performed according to the guidelines of the Declaration of Helsinki. In the ViSiElse graph (Fig. 5), the long action, entitled "intubation duration" allows us to see that midwives performed EI heterogeneously during neonatal resuscitation. Some midwives intubated early in their resuscitation efforts while others started after 4 min elapsed. ILCOR recommendations state that intubation should not occur during the first minute of life. The appropriate time for the insertion of the laryngoscope blade into the newborn's mouth is between 120 and 210 seconds, which was displayed by the green and black zones. ViSiElse allows a graphical inspection of the adequacy of the recommendations for medical procedures and provides a visual assessment of the performance of caregivers during training.

Online shopping behavior

Online shopping behavior is defined as the process in which consumers purchase items over the Internet. Comegys et al. (2006) described this process in a five step model: need recognition, information searches, evaluation, purchase decision, and post-purchase behavior. The authors compared online shopping behavior in 2002 and 2004/2005 in two countries (USA and Finland) and discovered that many factors influenced the buying process, including gender, age, education, and income (Jusoh & Ling, 2012; Wu, 2003). ViSiElse enabled the visualization of different groups of behavior and the first four steps of the online shopping behavior model used in Comegys et al. are modeled in a ViSiElse example (Fig. 6). The dataset is simulated for one hundred consumers divided into groups of 50 women in pink and 50 men in blue, allowing researchers to visually assess the differences in online shopping behavior between different categories of consumers. The ViSiElse graph also displays the summary statistics for each group. ViSiElse representations can be used to visualize any web navigation behavior.

Range of possible uses

ViSiElse was developed to meet the need to visualize raw behavioral data. However, the user-friendly package can be applied to all time data collected from a linear process, regardless of the research field. ViSiElse may be an asset in the field of cognitive ergonomics and the development of training programs or human-machine interaction, as well as in assembly lines to optimize linear processes and improve timing efficiency. ViSiElse can be used as visual feedback tool for data that are automatically extracted, as in healthcare simulations where a lot of software for automated data extraction exists. Simulation sessions often end with a debriefing between the examiners and the subject and, with ViSiElse, the debriefing could include an instantaneous visualization of the subject's performance.

Discussion

ViSiElse is a graphical tool developed to visualize raw data gathered from experimental observations of individual and/or group behavior over time. ViSiElse is a package of the open-source software R that can be applied to visualize any behavioral interactions between an organism and a process.

The package includes many features to provide a global overview of the data and can be used in the following ways:

- inspection of raw data;
- verification of the time adequacy to a procedure;
- global visualization to understand a behavior;
- apprehend learning processes or learning changes by using ViSiElse on repeated measures or sessions;
- compare group behavior.

With the inspection of the raw data, users can visualize the data distribution and identify outliers, which is especially useful when performing parametric tests that are sensitive to non-normal data and outliers. ViSiElse helps check statistical assumptions before running parametric tests.

Limitations

ViSiElse is currently limited to the visualization of raw data from procedures that can be linearized. Many complex procedures can be divided into processes that can be linearized, however, it is not possible to visualize events that involve multitasking or teamwork at this time. ViSiElse only offers descriptive support of raw data and must be integrated with complementary tools for a complete data analysis. However, ViSiElse is also able to extract patterns from graphs, run complete statistical analyses, and examine the quality of the action's execution. As ViSiElse uses raw time data, it can be associated with software that manages data extraction from video recorded sessions or software that directly provides time data.

ViSiElse's visualization on-screen is limited by the pixel pitch; the maximum discrimination capacity for long actions is limited to 725 individuals on a 21.5-inch screen with a resolution of 1920x1080 pixels and a pixel pitch of 0.248. However, on this screen, a maximum of 60 punctual actions can be plotted per graph without any limitation of the number of individuals.

Finally, ViSiElse only handles processes with time variations of the same scale. As the plotted result have a unique time axis with no time gap allowed, users are unable to observe time variations in seconds and hours. If a procedure involves actions that must be achieved within seconds and others within hours, we suggest splitting the process according to the time scale.

Future work

The features of the ViSiElse package will be expanded in the future; increasing the number of groups or the ability to plot nonlinear processes would help visualize more complex procedures involving parallel actions or interdisciplinary teams. The visualization of both quantitative and qualitative data will also be improved where qualitative data could be an indication of the goodness of the performance of the actions. Finally, color gradients, like those already used for punctual actions, may be added for long actions, which will remove the restriction on the number of individuals that can be plotted in a single graph.

In addition to improving the features of ViSiElse, future work will expand the ranges of data types allowed by creating an option to change the x-axis unit. Users could then visualize any quantitative variable. For example, concentrations or intensity could be visualized under different conditions. This improvement could extend the potential uses of ViSiElse. Similarly, an online interactive version of ViSiElse would broaden its availability and facilitate its use for any novice in data analysis.

Conclusion

ViSiElse is a package from R, the open-source software for statistical computing and data analysis. ViSiElse transforms a raw time data matrix into a comprehensible graph for immediate insight into the behavior of an individual or group. In a single one-page graph users are able to check their raw data, visualize time accuracy, compare groups, and analyze data distribution using summary statistics and tests. ViSiElse is accessible by ViSibook for the action characteristics (names, labels, types, order, delimitations and green and black zones), through the arguments of the visielse function for analysis features (time scale, groups, statistics), and through the arguments of the plot function for formatting options (labels size and color, adding a title, time interval size and unit). This graphical tool is suitable for use on every time-related process composed of actions that can be linearized. It was originally developed for use in the medical field but can be applied across all fields that use time data to analyze behavior. ViSiElse allows data reliability and transparency from an entire dataset to be assessed in a single view.

Acknowledgements

The authors thank Professors Jean-Bernard Gouyon and Simon Lorrain for their help in revising the manuscript and PeerJ language editing staff for revising the English.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). *Raincloud plots: a multi-platform tool for robust data visualization*. 48. https://doi.org/10.7287/peerj.preprints.27137v1

Almeida, A., & Azkune, G. (2018). Predicting Human Behaviour with Recurrent Neural Networks. *Applied Sciences*, *8*(2), 305. https://doi.org/10.3390/app8020305

Brewin, J., Tang, J., Dasgupta, P., Khan, M. S., Ahmed, K., Bello, F., … Jaye, P. (2015). Full immersion simulation: validation of a distributed simulation environment for technical and non-technical skills training in Urology. *BJU International*, *116*(1), 156–162. https://doi.org/10.1111/bju.12875

Comegys, C., Hannula, M., & Väisänen, J. (2006). Longitudinal comparison of Finnish and US online shopping behaviour among university students: The five-stage buying decision process. *Journal of Targeting, Measurement and Analysis for Marketing*, *14*(4), 336–356. https://doi.org/10.1057/palgrave.jt.5740193

Ellis, D. A., & Merdian, H. L. (2015). Thinking Outside the Box: Developing Dynamic Data Visualizations for Psychology with Shiny. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01782

Fosang, A. J., & Colbran, R. J. (2015). Transparency Is the Key to Quality. *Journal of Biological Chemistry*, *290*(50), 29692–29694. https://doi.org/10.1074/jbc.E115.000002

Garnier, E., Fouret, N., & Descoins, M. (2018). *Package ViSiElse. A Visualization Tool for Behavior Analysis.* Retrieved from https://cran.r-project.org/web/packages/ViSiElse/ViSiElse.pdf

Goedhart, J. (2018). *PlotsOfData - a web app for visualizing data together with its summaries*. 6. https://doi.org/10.1101/426767

Hertel, J. (2018). A Picture Tells 1000 Words (but Most Results Graphs Do Not). *Clinics in Sports Medicine*, *37*(3), 441–462. https://doi.org/10.1016/j.csm.2018.04.001

Jusoh, Z., & Ling, G. H. (2012). FACTORS INFLUENCING CONSUMERS' ATTITUDE TOWARDS E-COMMERCE PURCHASES THROUGH ONLINE SHOPPING. *International Journal of Humanities and Social Science*, *2*(4), 8.

Kalaniti, K., & Campbell, D. M. (2015). Simulation-based medical education: Time for a pedagogical shift. *Indian Pediatrics*, *52*(1), 41–45. https://doi.org/10.1007/s13312-015-0565-6

Larson-Hall, J. (2017). Moving Beyond the Bar Plot and the Line Graph to Create Informative and Attractive Graphics1. *The Modern Language Journal*, *101*(1), 244–270. https://doi.org/10.1111/modl.12386

Pastore, M., Lionetti, F., & Altoè, G. (2017). When One Shape Does Not Fit All: A Commentary Essay on the Use of Graphs in Psychological Research. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01666

Prager, E. M., Chambers, K. E., Plotkin, J. L., McArthur, D. L., Bandrowski, A. E., Bansal, N., … Graf, C. (2018). Improving transparency and scientific rigor in academic publishing: PRAGER et al. *Journal of Neuroscience Research*. https://doi.org/10.1002/jnr.24340

R Core Team. (2018). R: A Language and Environment for Statistical Computing. *Foundation for Statistical Computing, Vienna*. Retrieved from https://www.R-project.org

Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, 5.

van Kasteren, T., Noulas, A., Englebienne, G., & Kröse, B. (2008). Accurate activity recognition in a home setting. *Proceedings of the 10th International Conference on Ubiquitous Computing - UbiComp '08*, 1. https://doi.org/10.1145/1409635.1409637

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*, *13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128

Weissgerber, T. L., Savic, M., Winham, S. J., Stanisavljevic, D., Garovic, V. D., & Milic, N. M. (2017). Data visualization, bar naked: A free tool for creating interactive graphics. *Journal of Biological Chemistry*, *292*(50), 20592–20598. https://doi.org/10.1074/jbc.RA117.000147

Wu, S. (2003). The relationship between consumer characteristics and attitude toward online shopping. *Marketing Intelligence & Planning*, *21*(1), 37–44. https://doi.org/10.1108/02634500310458135

Wyckoff, M. H., Aziz, K., Escobedo, M. B., Kapadia, V. S., Kattwinkel, J., Perlman, J. M., … Zaichkin, J. G. (2015). *Part 13: Neonatal Resuscitation*. 18.

Ziv, A., Wolpe, P. R., Small, S. D., & Glick, S. (2006). Simulation-Based Medical Education: An Ethical Imperative: *Simulation In Healthcare: The Journal of the Society for Simulation in Healthcare*, *1*(4), 252–256. https://doi.org/10.1097/01.SIH.0000242724.08501.63

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019

PeerJ Preprints NOT PEER-REVIEWED

Figure 1

Actions of a typical day represented with a ViSiElse.

This figure shows ViSiElse's representation of the everyday life tasks over time (8 punctual and 4 long actions) based on a simulated dataset of a hundred participants. ViSiElse's legend is divided into two parts:

- The left side is the legend for punctual actions. The first column displays the gradient of colors proportional to the number of participants represented in each time interval of 30 min. The second column shows the time constraints and summary statistics. Time constraints are set at the start and the stop of the working hours with a black zone for inadequate arrival (after 10 a.m.) and departure (before 4 p.m.). Additionally, time constraints are set on the time to pick up the kids with one green zone for the adequate period (from 4 to 5 p.m.) and two black zones for inadequate time. Summary statistics for punctual actions are median, first and third quartile (line and dots).

- The right side is the legend for long actions. The first row represents the time constraints. The long action "lunch break" should not last more than 30 min, the inadequate duration is displayed by a darker blue color. The second row shows the summary statistics for long actions symbolized by a line proportional to the median duration of the actions.

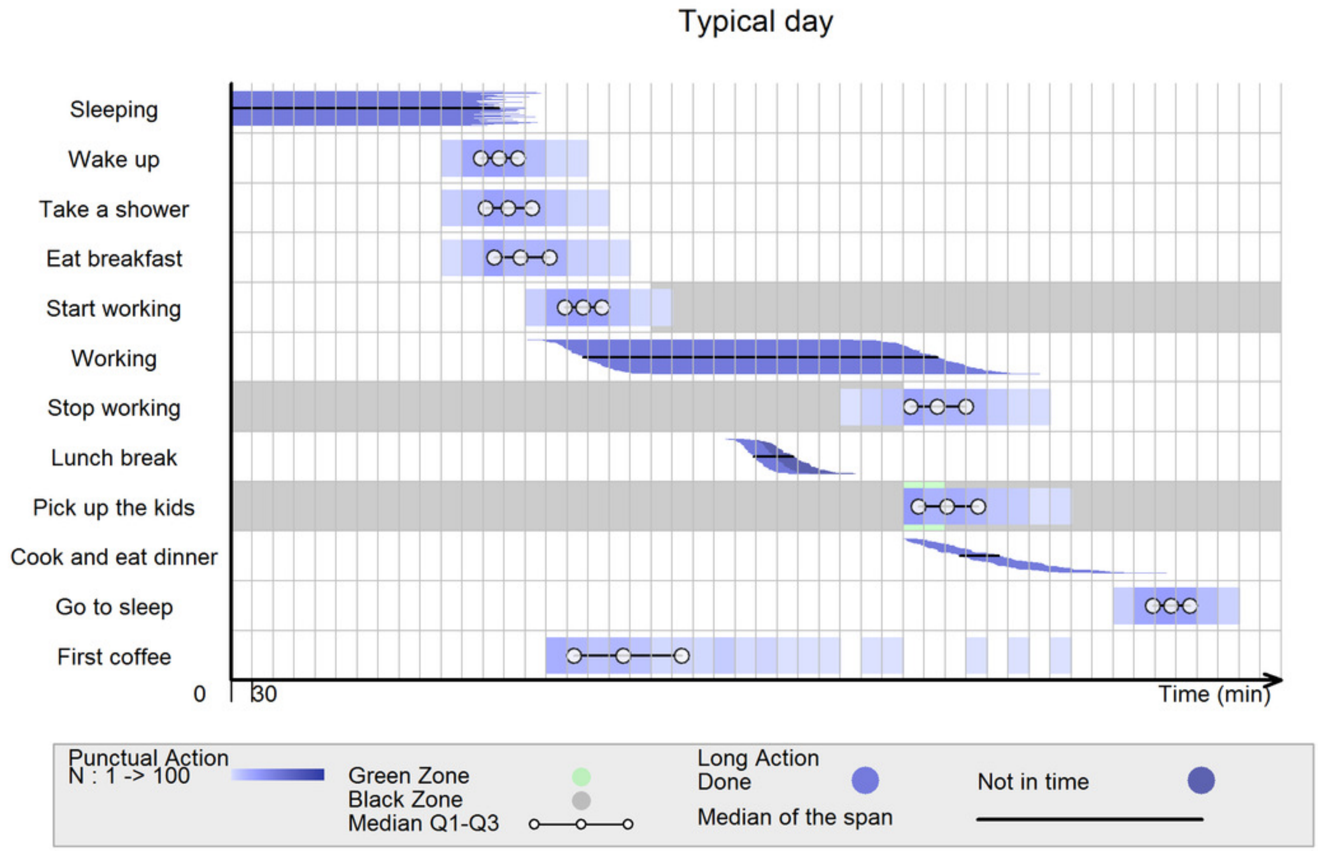
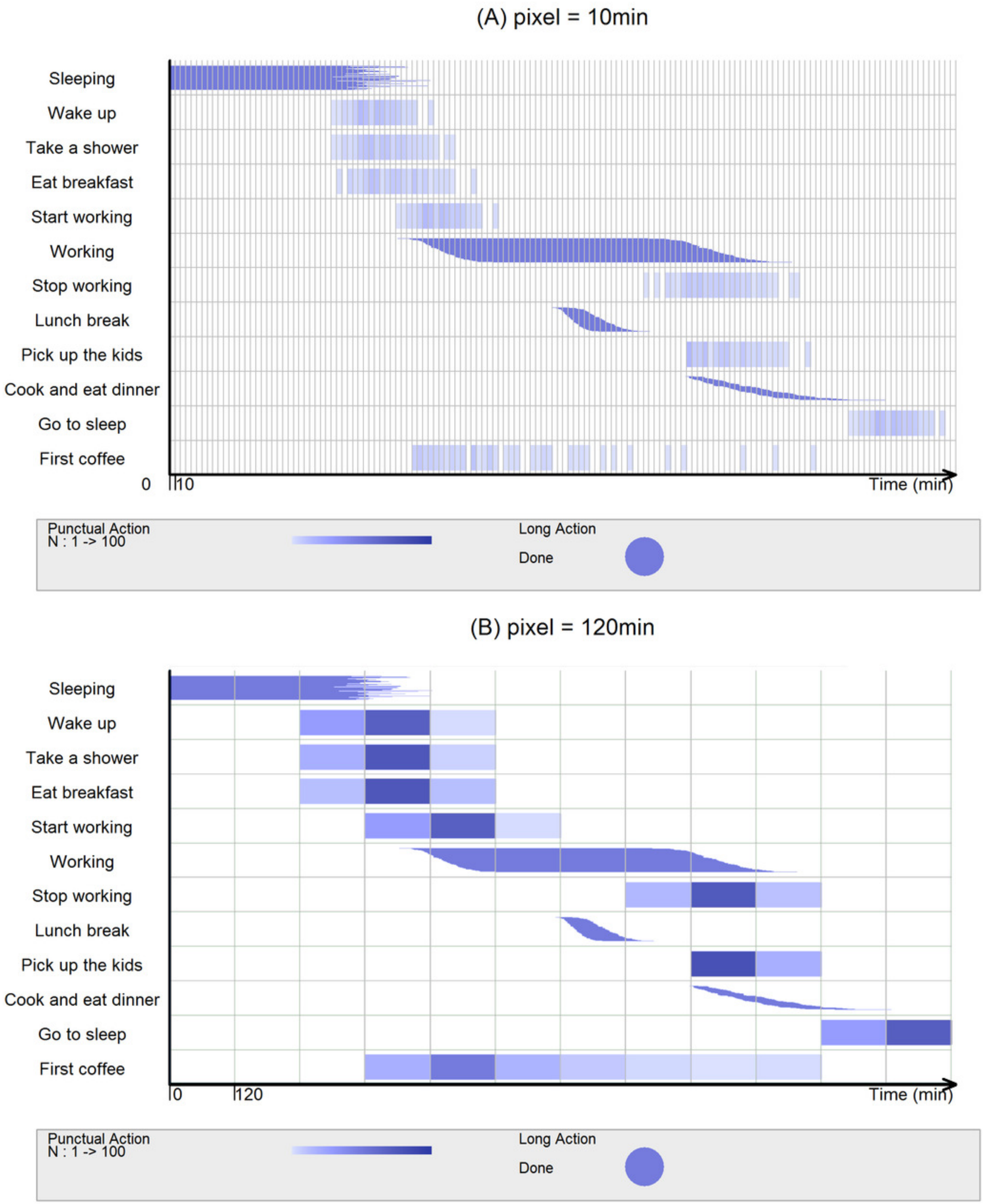


Figure 2

Graphical consequences of the modification of the pixel parameter

VisiElse pixel is a key parameter linked to the behavior observed duration. It should be carefully set. In panel *(A) ViSiElse graph with pixels = 10 min.* The too-short pixel duration made participants data not enough aggregated to allow a clear visualization*. (B) ViSiElse graph with pixels = 120 min* participants were too much aggregated resulting in a loss of information about the statistical distribution of the participants over the actions.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019

Figure 3

ViSiElse graph with three different methods to plot groups.

The three graphs show the typical day actions for two groups: participants who employ a babysitter are displayed in blue while participants who do not are in pink. For each action, mean and standard deviation are presented. *(A) The “cut” method* is used: groups1 and 2 are one under the other. Each group has its own statistical indicators. *(B) The “join” method* is used: groups are mixed together but differentiable by colors. Statistical indicators are calculated for all the individuals and not per group. *(C) The “within” method* is used: groups are plotted together in blue and group1 is plotted again in pink. The first statistical indicator is for the global data and the second is for the repeated group.

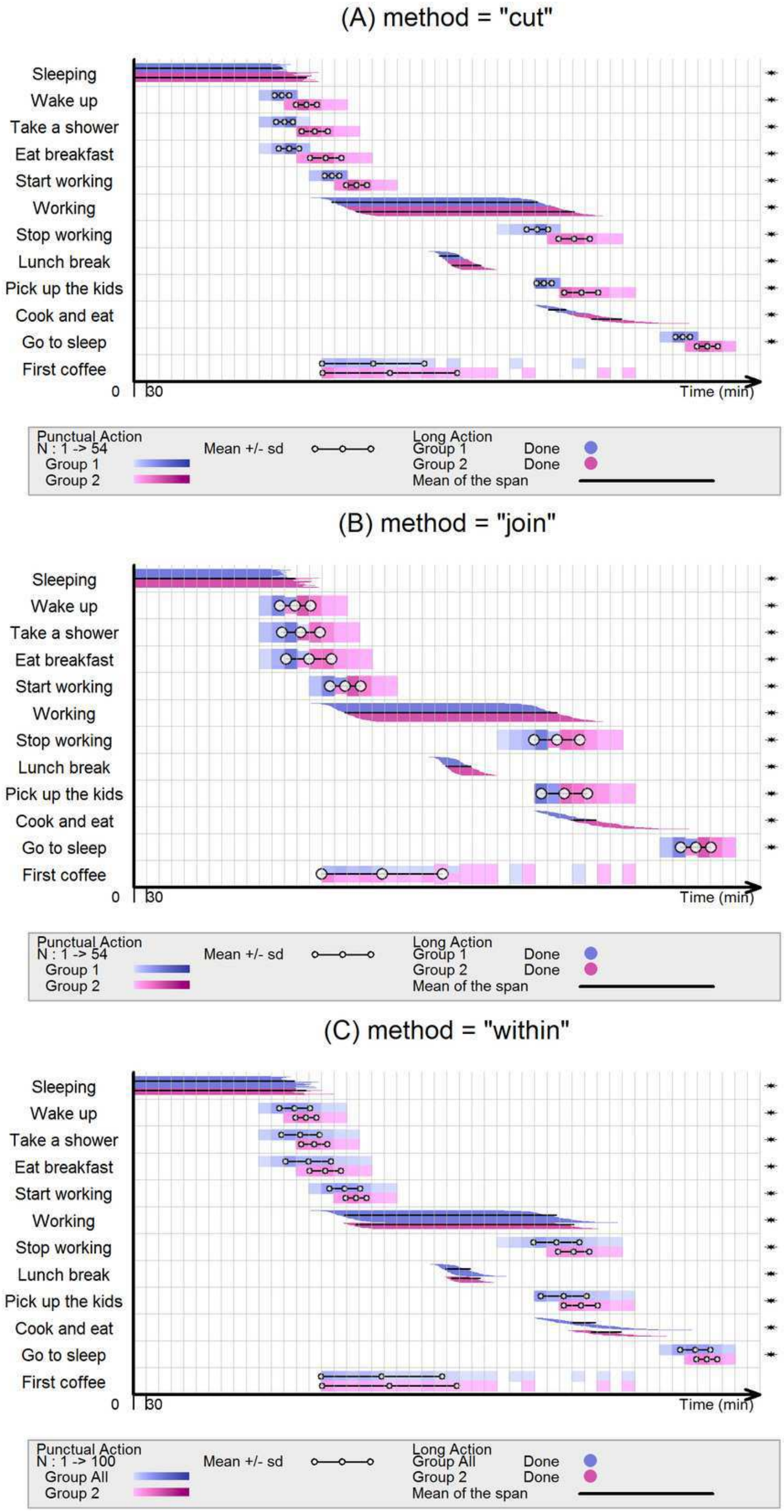

(A) method = "cut"
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat
Go to sleep
First coffee
0
30
Time (min)
Punctual Action
N : 1 -> 54
Group 1
Group 2
Mean +/- sd
Long Action
Group 1 Done
Group 2 Done
Mean of the span
(B) method = "join"
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat
Go to sleep
First coffee
0
30
Time (min)
Punctual Action
N : 1 -> 54
Group 1
Group 2
Mean +/- sd
Long Action
Group 1 Done
Group 2 Done
Mean of the span
(C) method = "within"
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat
Go to sleep
First coffee
0
30
Time (min)
Punctual Action
N : 1 -> 100
Group All
Group 2
Mean +/- sd
Long Action
Group All Done
Group 2 Done
Mean of the span

Figure 4

Examples of other raw data visualization tools.

The three graphs represent the same typical day dataset with different visualization methods. *(A) Scatter plot.* Each action is plotted separately. Advantage: easy to use; drawback: cannot visualize the entire process at once or the order of the actions. *(B) Violin + scatter plot.* Each line represents an action. Advantage: visualization of the distribution; drawback: a limited number of actions plotted simultaneously. *(C) Heatmap.* Each line represents an action. Advantage: compact visualization; drawback: no punctual/long actions distinction and no group distinction.

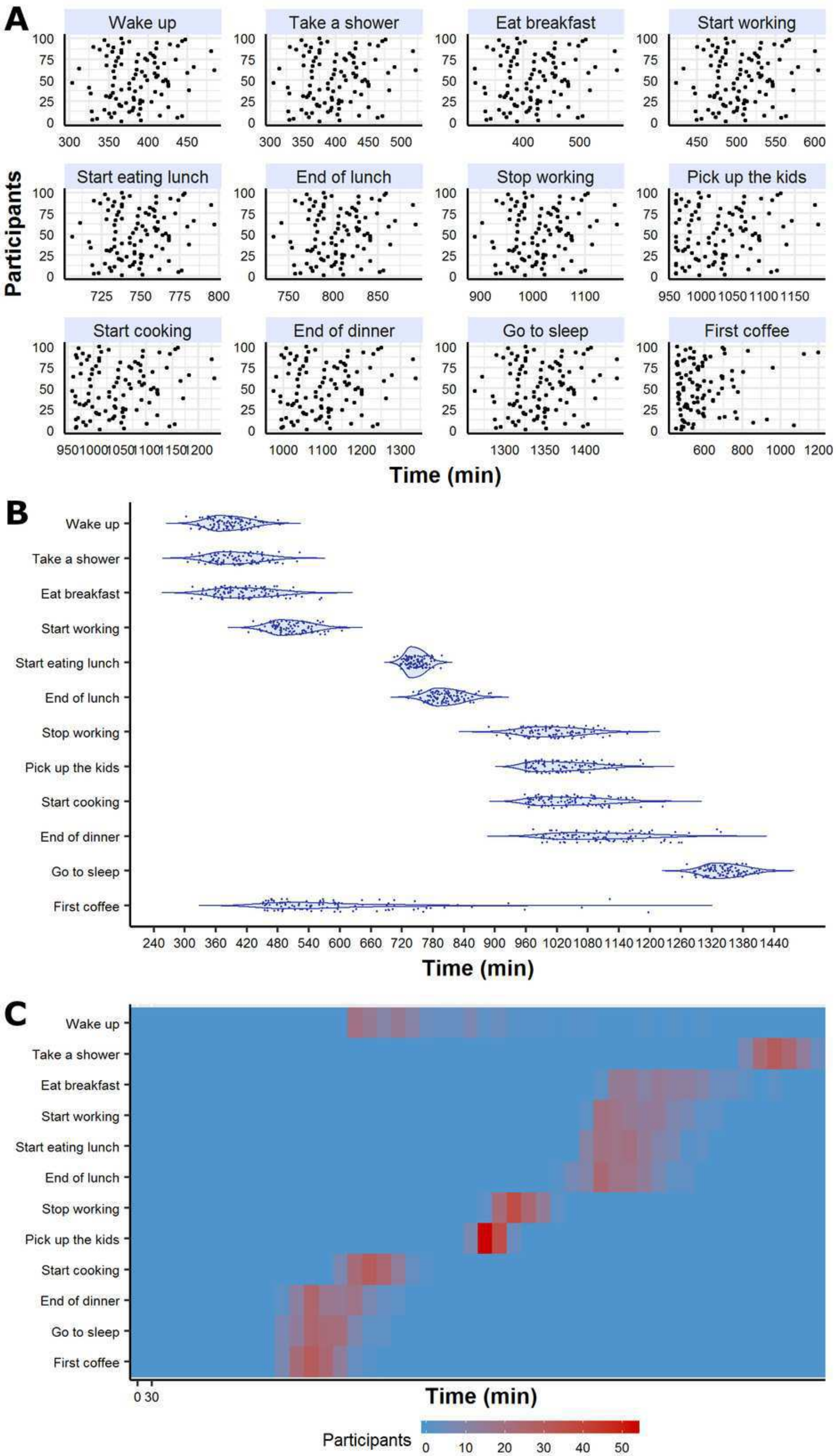
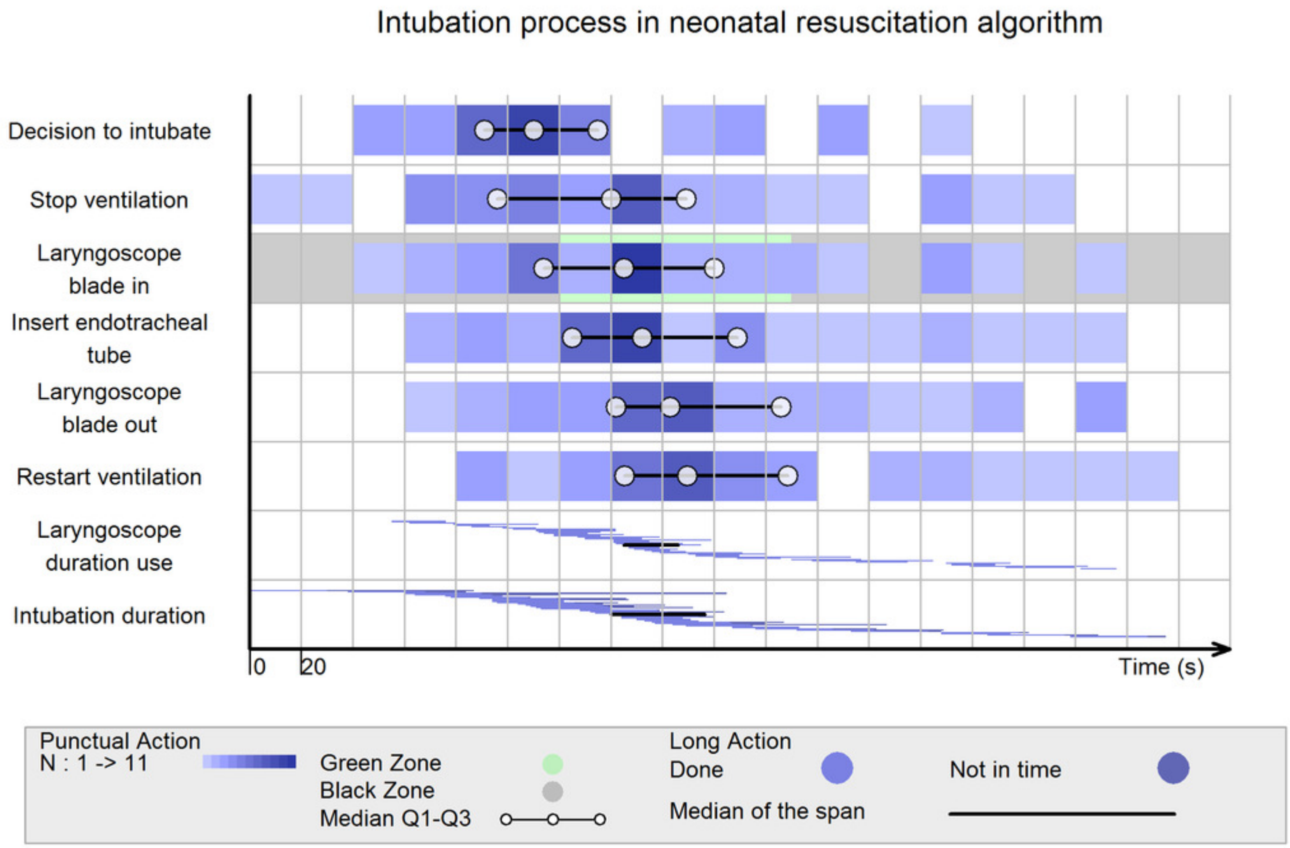

A
Wake up
Take a shower
Eat breakfast
Start working
Start eating lunch
End of lunch
Stop working
Pick up the kids
Start cooking
End of dinner
Go to sleep
First coffee
Participants
Time (min)
B
Time (min)
C
Time (min)
Participants

Figure 5

ViSiElse graph of the orotracheal intubation process during simulated neonatal resuscitation.

This figure shows ViSiElse's representation of the orotracheal intubation process (8 actions) during a simulated neonatal resuscitation realized by 37 participants. Statistical indicators for punctual actions are interquartile range (line and dots) and for long actions median of the duration (line). Time constraints are set on the insertion of the laryngoscope blade in the newborn's mouth with one green zone for the adequate period (between 120 and 210 seconds) and two black zones for inadequate time. An additional time constraint is set on the duration of the intubation process shown by a darker blue color when the process lasts more than 30 seconds ("Not in time").



PeerJ Preprints NOT PEER-REVIEWED

Figure 6

ViSiElse graph of a simulated dataset of online shopping behavior.

This figure shows ViSiElse's representation of online shopping behavior (4 actions). The simulated dataset is based on the five stages buying decision process model described in 2006 by Comegys, Hannula, & Väisänen (the post-purchase behavior stage was omitted). Data are separated into two groups: 50 men in blue and 50 women in pink. Statistical indicators for punctual actions are mean and standard deviation (line and dots) and for long actions mean of the duration (line).

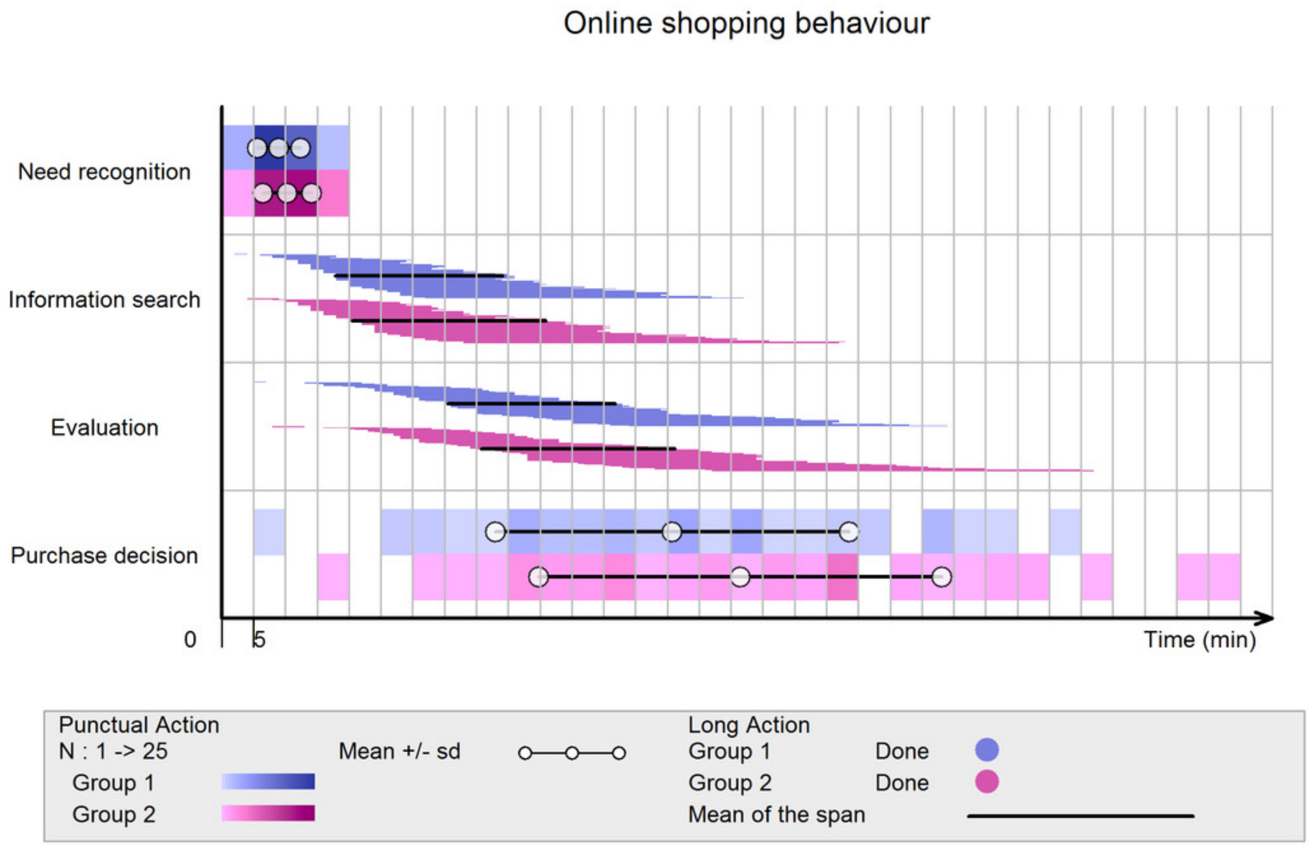


Table 1(on next page)

First five rows of the typical day dataset

The first column is the subject ID and the following columns are the timestamps of the actions (in minutes). The timestamps are the duration elapse from the starting point to the action. In the typical day dataset, the starting point is midnight (“start_sleep”) and, for example, the timestamp of the action “wake up” is the duration between midnight and the waking moment in minutes. Subject 1 woke up at 6h06 so the timestamp is 366 min.

id	start_sleep	stop_sleep	wake_up	shower	breakfast	start_work	start_lunch	stop_lunch	stop_work	pickup_kids	start_cook	stop_cook	go_sleep	first_coffee
1	0	366	366	375	389	486	738	789	985	997	1011	1059	1326	479
2	0	391	391	406	426	511	751	811	1022	1037	1057	1118	1351	451
3	0	329	329	329	334	449	720	757	929	960	965	995	1289	535
4	0	335	335	336	342	455	723	763	938	960	966	999	1295	489
5	0	437	437	464	496	557	774	852	1091	1112	1144	1228	1397	481

1

Table 2 (on next page)

Graph comparison between Scatter plot, Violin + Scatter plot, Heatmap and ViSiElse graph

Four raw data visualization tools are evaluated based on the characteristics that best represent time series of actions. The ease of use refers to the complexity required to create the graph: combination of graphs (violin + scatter plot), data manipulation (heatmap) or additional information (ViSiElse).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019

	Scatter plot	**Violin + Scatter plot**	**Heatmap**	**ViSiElse**
Raw data	X	X	X	X
No data manipulation	X	X		X
Process visualization		X	X	X
High-dimensional dataset			X	X
Distribution visualization		X	X	X
Punctual actions	X	X	X	X
Long actions				X
Statistical indicators	Mean	IQR		Mean + SD *or* IQR
Group distinction	X	X		X
Time accuracy				X
Ease of use	Easy	Medium	Medium	Medium

1

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v3 | CC BY 4.0 Open Access | rec: 25 Nov 2019, publ: 25 Nov 2019