ViSiElse: An innovative visualization R package to ensure behavioral raw data reliability and transparency

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

Background. In recent years, the scientific community encouraged the use of raw data graphs to improve the reliability and transparency of the results presented in papers. However, methods to visualize raw data are limited to one variable per graph and/or only small samples. In behavioral science as in many other fields, multiple variables need to be plotted together to allow insights of the behavior or the process observations. In this paper, we present ViSiElse, an R-package offering a new approach in raw data visualization.

Methods. ViSiElse was developed with the open-source software R to provide a solution for the complete visualization of the raw time data.

Results. ViSiElse grants a global overview of a process by combining the visualization of multiple actions timestamps for all participants in a single graph. Individuals and/or group behavior can easily be assessed. Supplementary features allow users to further inspect their data by adding statistical indicators and/or time constraints. ViSiElse provides a global visualization of actions acquired from timestamps in any quantifiable observations.

ViSiElse: An innovative visualization R package to ensure behavioral raw data reliability and transparency

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

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

Corresponding Author:
Elodie Garnier[1]
Centre d'Étude Périnatales de l'Océan Indien (CEPOI, EA 7388), CHU de la Réunion, 97410 Saint Pierre, France.
Email address: e.garnier30@gmail.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 2019

Abstract

Background. In recent years, the scientific community encouraged the use of raw data graphs to improve the reliability and transparency of the results presented in papers. However, methods to visualize raw data are limited to one variable per graph and/or only small samples. In behavioral science as in many other fields, multiple variables need to be plotted together to allow insights of the behavior or the process observations. In this paper, we present ViSiElse, an R-package offering a new approach in raw data visualization.

Methods. ViSiElse was developed with the open-source software R to provide a solution for the complete visualization of the raw time data.

Results. ViSiElse grants a global overview of a process by combining the visualization of multiple actions timestamps for all participants in a single graph. Individuals and/or group behavior can easily be assessed. Supplementary features allow users to further inspect their data by adding statistical indicators and/or time constraints. ViSiElse provides a global visualization of actions acquired from timestamps in any quantifiable observations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 2019

Introduction

Time data are widely used in multiple research fields including economics, biology, medicine or social sciences. Time data are temporal observations acquired from different sources like video recorded experiments, sensors, web navigation or direct measurements. 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, "non-raw time data" refers to processed or summarized time data.

Behavioral science focuses on the development of behavioral knowledge through experimental observations where behavior can be defined as a series of actions. The analysis of the action timestamps allows researchers to determine if the observed behavior is appropriate. The right behavior is the series of actions leading to the achievement of a goal as the successful completion of a defined process.

To capture the right behavior, data must be explored. Raw time data stored in large tables are so difficult to get to grips with. Direct interpretation of raw data tables with lots of actions and lots of participants is impossible. Graphical representations are, on the contrary, effortless, instantaneous and easy to understand. The ideal visualization of behavioral data is a graphic tool that simultaneously plots the raw time data for each action and for all participants. There are numerous methods to visualize data but tools to visualize raw data with several variables in the same graph are limited.

The scientific community recently pointed out the value of raw data presentation in publications (Fosang & Colbran, 2015; Prager et al., 2018; Rousselet, Foxe, & Bolam, 2016) and demonstrated how graphic methods that summarize data can be misleading and even suggest different conclusions from the reality of the data distribution (Weissgerber, Milic, Winham, & Garovic, 2015).

With the growing idea of open science, investigators are now encouraged to be transparent. Plotting raw data not only increases clarity but also makes results more understandable and more reliable. Recommendations include choosing the relevant plot according to the data and favoring graphics that present all of the data as well as their structure like scatter plots, violin plot, beeswarm or pirate plot over graphics that only present a summary such as bar plots or line plots (Hertel, 2018; Larson-Hall, 2017; Pastore, Lionetti, & Altoè, 2017). Allen et al. (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2018) introduced raincloud plots, an easy to use multi-platform tool that combines visual representations. Raincloud provides a complete overview of the data and makes robust and transparent plots. Other researchers turned to interactive visualizations (Ellis & Merdian, 2015; Goedhart, 2018; Weissgerber et al., 2017), letting readers explore the dataset with customizable graphs and additional features for statistical indicators. All of those solutions successfully answered the need for reliability and transparency in publications. However, the visualization of raw data is limited to one variable per graph and/or small samples. To our knowledge, none proposed an innovative raw time data visualization tool displaying an entire process for large samples of participants in a single graph.

To answer this limitation, we developed ViSiElse. ViSiElse is a package 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/web/packages/ViSiElse/index.html. ViSiElse is a graphical tool created to visualize and to provide a global insight of individuals and/or group actions that define behavior. ViSiElse allows visualization of raw time data extracted from any experimental observations. Options for the package include additional graphical information as statistical indicators (mean, standard deviation, quartiles or statistical tests) and also, for each action, green or black zones providing visual information about 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.

In this paper, we provide a step-by-step presentation of ViSiElse main features. We describe how to set up data to create a ViSiElse plot, how to custom this plot to get a clear view of the individuals' actions, and how to add time limits or statistical indicators. As a supporting example, we use a simulated dataset of the actions performed on a typical day. Then we compare ViSiElse with other raw data visualization tools. Finally, we describe different applications and discuss the range of ViSiElse possible uses.

The online supplementary material contains complete data, and R code to reproduce the following presented results (https://github.com/CEPOI/ViSiElse/tree/master/Example).

Methods

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

Creation of the dataset and ViSibook objects

To build the dataset correctly, researchers have to a priori translate a behavior or a procedure into a linear process of actions. This transformation has been described in Almeida & Azkune (Almeida & Azkune, 2018) where the authors decomposed behavior into activities and then activities into actions. For ViSiElse, three elements must be defined: the list of actions, their type, and their order. The fundamental question to answer is: "What are the elementary actions composing the behavior?". An elementary action is an action that cannot be divided into shorter actions regarding the time scale. For example, the typical day tasks can be divided into the following elementary actions: sleep, wake up, take a 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. used a similar decomposition of everyday activity but their study only focused on actions achieved at home (van Kasteren, Noulas, Englebienne, & Kröse, 2008).

Once the list of elementary action is established, they should be classified as punctual or long. A punctual action is an action with no duration, or not lasting long enough to be measured regarding the time scale of the studied behavior. A long action is an action having duration defined by two punctual actions, one for its beginning, and one for its ending. For example, the action “sleep” is long—it lasts at least a few hours—while the action “wake up” is punctual—it usually only takes seconds or a few minutes. 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. If two actions are happening simultaneously, we suggest randomly ranking them. Not attributing a rank is only relevant when the interest of a punctual action is purely to define a long one. Non-ranked action will not be plotted.

For the typical day example, the final list of actions 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 correspond to the completion times of each punctual action for all participants or groups of participants. The dataset is the organized raw data. The first column of the dataset identifies individuals. Other columns represent punctual actions including the punctual actions that delimit the long actions. Lines contain the timestamps of each action for each individual, without any calculation or transformation. An example of the appropriate data structure is the typical day dataset (ViSiElse_typDay_data.csv). This dataset is an R-simulated dataset and is available online at https://github.com/CEPOI/ViSiElse/tree/master/Example.

Building the ViSibook. While the dataset contains the raw time data of the studied behavior, the ViSibook contains the structure of the behavioral process. Mainly, it is a table consisting of the definition of every action. The minimum structure for a ViSibook must give for each action its name, its label, its type (punctual or long), its order in the behavioral process, and for long actions only, the name of the two punctual actions delimiting its beginning and its ending. Optionally, the ViSibook can include time constraints (green and/or black zones) which provide visual information about the time accuracy of the realized actions. To create a ViSibook, users can define or import a table as long as they carefully keep the names and the order of the ViSibook's columns—see ViSiElse CRAN documentation (Garnier et al., 2018). The ViSibook is an optional argument of the main function “visielse”. “visielse” function generates a ViSiElse graph of the studied behavior, according to the dataset and the ViSibook. When it is not specified, “visielse” computes a default ViSibook from the dataset. In this case, the process order is given by the dataset column names and all actions are set to punctual. At any time, the ViSibook can be extracted from an execution of the “visielse” function. Users can then modify any information saved in the ViSiBook to adjust the plotted behavior. For example, on ViSiElse graph, the names of the displayed actions on the y-axis are the labels set in the ViSibook. By default, those labels are the variable names defined by the dataset column names. Usually, variable names are a single short word lacking clarity to describe an action. Therefore it is often preferred to use a more explicit description as displayed labels. Through the ViSibook, users can change the action labels to improve clarity in the graph.

Visualization of Raw Time Data With ViSiElse

ViSiElse gives, in a one-page single graph, an overview of the distribution of the timestamps of sequential actions for large samples of participants. This innovative visualization facilitates the comprehension of individuals and/or groups’ behavior based on the profile of the data distribution. Additionally, ViSiElse eases the identification of outliers or abnormal behaviors, i.e. a behavior not complying with good practice recommendations.

Creating the first plot. Simply running the “visielse” function with a dataset and an optional ViSibook as arguments will create and display the ViSiElse graph. Fig 1 shows the simulated typical day dataset for one hundred participants. On the graph, actions are organized one under the other (y-axis) and their executions are distributed along the time axis (x-axis). For punctual actions, a drawn rectangle means that at least one individual has done the action in the interval of time. The size of the time interval is set by the breaks of the time axis. In Fig 1, the breaks are set every 30 min from midnight to midnight. The color's intensity of the rectangles is proportional to the number of individuals who realized the action during the time interval. For long actions, individuals are represented by lines whose sizes are proportional to the duration of the action. Lines are chronologically sorted by the action starting time.

To access and adjust the formatting options of the graph, the ViSiElse object should be plotted using the R basic "plot" function. Formatting options include many properties like changing the label's size and color, adding a title or modifying the time interval size and unit—set to 10 seconds by default. Modifying the size of the time interval through the plot function with the "scal.unit.tps" parameter will only change the breaks of the time axis and not the size of the time interval used to calculate the color's intensity of the punctual actions (see the "pixel" parameter below). Every formatting option is accessible through the "plot" function while the package features (group comparison, time constrains, statistical indicators) must be defined in the "visielse" function.

Adjusting the time interval with the pixel parameter. For punctual actions, the "pixel" parameter represents the time precision i.e. the time limit to which one subject is moved from a 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 at least inferior—as the pixel parameter defined in the ViSiElse object. Indeed, the "plot" function changes the formatting option of the ViSiElse graph so the two parameters should be coherent.

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

Analysis of Raw Time Data With ViSiElse

ViSiElse offers many features to analyze behavioral raw time data. For example, users may define groups, time constraints or statistical indicators to complete their graph. ViSiElse helps with both the inspection and the interpretation of raw time data in a single graph.

Compare group behavior. ViSiElse enables the differentiation of two subsets of participants through color distinction. Distinguishing experimental groups of participants is useful to identify different patterns of behavior. In the typical day dataset, we created two groups: people who employ a babysitter (in blue) and people who do not (in pink). To display groups with 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 to plot groups:

- The “cut” method where each group is represented one under the other with different colors (see Fig 3.A). This representation is useful to compare groups as data are completely dissociated.

- The "join" method where groups are spatially mixed but are differentiated by distinct colors (see Fig 3.B). 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 behavior against the global population. As the ViSiElse package only allows two colors of distinction, the "within" method is the most suitable option for data containing more than two groups. Users can confront each group, one after the other, to the global population and identify their differences.

Set time constraints. Behavior may sometimes be constrained by external guidelines where actions must respect an order and a timing. When it occurs, punctual actions should either be achieved during a specific period or not be executed before or after a specific time point. Likewise, long actions should not exceed a specific duration or continue after a specific time point. ViSiElse uses green and black zones to help visualize such time boundaries. Green zones represent time obligation i.e. when actions should be achieved. Black zones set time interdiction i.e. when actions should not occur. Setting visual time constraints shows whether or not the individuals' behavior is done during the appropriate time zone. For each punctual action, users can define one green zone and two black zones (one before and one after the expected execution times). To create those time zones, users only have to define their delimitations: 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. They are automatically plotted when running the "visielse" function. ViSiElse also allows the repetition of green zones when a punctual action can be achieved in different time zones. For this option, users define in the ViSibook the time points of the first green zone, and then, the time interval between each green zone. For long action, ViSiElse only offers time interdiction—black zones. As long actions are in fact duration, there are two ways of restricting them: a deadline not to cross or duration not to exceed. To define the time constraints of long actions, users must define the time points and the appropriate restriction method in the ViSibook.

In a typical day, actions are restrained by external rules. For example, the working hours are delimited. Usually, people should be at work and start working before 10 a.m. and they cannot leave work before 4 p.m. In addition, they only have a 30 min break for lunch. Finally, schools often end at 4 p.m. and close at 5 p.m. so people must pick up their kids during this one-hour interval. Therefore, in Fig 1, time constraints are set on multiple actions to assess if they are done in the appropriate time zones. The punctual actions “start working” and “stop working” each has a black zone for respectively inadequate arrival time (after 10 a.m.) and inadequate departure time (before 4 p.m.). The punctual action “pick up the kids” have one green zone for the adequate period (from 4 to 5 p.m.) and two black zones for inadequate time (outside the one-hour interval). The long action “lunch break” has a 30 min-duration limitation displayed by a darker blue color.

Analyze with statistical indicators. To complete the ViSiElse’s graph and analyze the behavioral data tendency, statistical indicators can be added. Users can choose between plotting the mean and standard deviation (see Fig 3) or the median with the first and third quartiles (see Fig 1). Moreover, if the data contains groups, and statistical indicators are defined, ViSiElse computes a statistical test to compare the time data between the two groups (see Fig 3.A-B). If the statistical indicator is set to mean and standard deviation, ViSiElse runs a Wilcoxon test; if the statistical indicator is set to median and quartiles, ViSiElse runs a Mood’s two-sample test. An asterisk appears on the right side of the graph if the statistical test is significant with a 0.01 alpha risk (default value). The alpha risk could be manually set. For example, in Fig 3.A-B, it was set to 0.05 resulting in a significant test for all actions except for the punctual action “first coffee”. 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. For a complete analysis, ViSiElse should be associated with other analytic tools.

Limitation of usual Raw Data Visualization Tools

A lot of graphical tools are available to plot and visualize data. We selected three of them that are the most suitable to time raw data. Table 1 summarizes the main characteristics of each raw data visualization method that we described.

To visualize raw data, we commonly use scatter plots. This visualization method is easy to use and perfect for a small number of variables. However, for a highly dimensional dataset, users need to display all variables one by one. For example, Fig 4.A shows the 12 graphs required to see the punctual actions of the typical day dataset. Interpreting the scatter plots is difficult 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.

A solution is to combine all the scatter plots into a single graph by using the violin plot representation. For example, Fig 4.B displays the same 12 punctual actions with both violin and scatter plots. The dots are the raw data while the violin shape shows the data distribution. The violin plots are usually presented vertically. Here we reversed the x and y-axis to keep the time axis horizontal. This visualization provides the global overview of the process. Users can add boxplots to get the data tendency. They can also display as many groups as required. The combination of violin and scatter plots is useful for medium-size datasets. However, if either the number of variables or the number of participants increases, the dots would be too much aggregated to allow interpretation.

For large and highly dimensional datasets, heatmaps are very efficient tools. Fig 4.C shows the heatmap of the typical day dataset. Similarly to ViSiElse graphs, heatmaps can display an unlimited number of participants and a large number of actions as it uses a gradient of color. With this visualization method, users can see the global process, the order of the actions over time and both the raw data and the data distribution. There are still some drawbacks. Heatmaps do not provide statistical indicators or group distinction. Additionally, heatmaps —and also violin and scatter plots— only allow punctual actions. Therefore long actions can only be displayed by their start time and end time. This is a major limitation when the duration of action matters. If we only see the start and end time of the action, we would not be able to visualize its duration. Indeed, when punctual actions are plotted, individuals are pulled together so we cannot link a start time to its end time: some might start early and finish early, others might start early but finish late or start late and finish late.

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 increasingly learned using high-fidelity simulation to avoid errors and give optimal care to the patients without any associated risk due to the learning process (Brewin et al., 2015; Kalaniti &

Campbell, 2015; Ziv, Wolpe, Small, & Glick, 2006). For example, midwife students are trained to the neonatal resuscitation procedure which includes endotracheal intubation (EI). EI is the process of inserting a tube through the mouth and then into the airway in order to ventilate. By restoring airway patency, EI is a lifesaving procedure and it has to be readily available to all patients whose ventilation is compromised in emergency or anesthesia context. 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, as most of the medical procedures, follows guidelines set by local or international committees. For the neonatal resuscitation procedure, the International Liaison Committee on Resuscitation (ILCOR) sets the guidelines that caregivers should follow, including those for EI (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) investigating the neonatal resuscitation training of midwives using 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), thanks to the long action "intubation duration", we see that midwives performed the EI heterogeneously during the neonatal resuscitation. Some midwives intubate early in the resuscitation while others only start after 4 min. ILCOR recommendations state that the intubation should not occur during the first minute of life. An appropriate time for the insertion of the laryngoscope blade in the newborn's mouth is between 120 and 210 seconds which is displayed by the green and black zones. For medical procedures, ViSiElse allows a graphical inspection of the adequacy to the recommendations. For training processes, ViSiElse provides a visual assessment of the performance of caregivers.

Online Shopping Behavior

Online shopping behavior is a buying process where consumers purchase items over the internet. In 2006, Comegys et al described this process in a 5 steps model: need recognition, information search, evaluation, purchase decision, and post-purchase behavior (Comegys, Hannula, & Väisänen, 2006). In their study, the authors compared online shopping behavior in 2002 and 2004/2005 for two countries (USA and Finland). Many factors influencing the buying process have been studied such as gender, age, education or income (Jusoh & Ling, 2012; Wu, 2003). ViSiElse enables the visualization of different groups of behavior. Fig 6 shows the ViSiElse graph of the first four steps of the online shopping behavior model used in Comegys et al. The dataset is a simulated dataset for one hundred consumers divided into two groups: 50 women in pink and 50 men in blue. With ViSiElse, researchers can visually assess the differences in online shopping behavior between different categories of consumers. ViSiElse graph also displays the statistical indicators for each group. Fig 6 is an example of ViSiElse application for online shopping data but ViSiElse representation can be used to visualize any web navigation behavior.

Range of Possible Uses

ViSiElse was developed as a solution to the lack of visualization tools for behavioral raw data, but the user-friendly package can be used for all time data collected from a linearized process regardless of the research field. For example, cognitive ergonomics, including the development of training programs or human-machine interaction, are study fields where ViSiElse can be an asset. ViSiElse can also be useful in assembly lines where processes are easily linearized and where timing is crucial. For example, ViSiElse could improve the time efficiency in an assembly line by easily identifying which steps of the process can be optimized. In the case of data automatically extracted, ViSiElse could also be used as visual feedback tool. For example, in healthcare simulation where a lot of software for automated data extraction exists, the simulation session often ends with a debriefing between the examiners and the subject. With ViSiElse, the debriefing could be upgraded with an instantaneous visualization of the subject's performance.

Discussion

We presented ViSiElse, a graphical tool conceived to visualize raw data from experimental observations of individuals and/or groups' behavior. 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 different 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 groups' behavior

With the inspection of the raw data, users can also visualize the data distribution and spot outliers. Those two features are 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

Currently, ViSiElse is limited to the visualization of raw data solely from procedures that can be linearized. Many complex procedures can be divided into short ones that can be linearized but visualizing multitasking or teamwork is currently not possible. ViSiElse's features only offer descriptive support of raw data. To provide complete data analysis, it should be integrated with complementary tools. Extracting patterns from the graph, running complete statistical analysis or examining the quality of the action's execution are many additional investigations that can be done. As ViSiElse uses raw time data, it can be associated with software that manages data extraction from video recorded sessions or software that directly provide time data like serious games.

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

Finally, ViSiElse only handles processes with time variations that are on the same scale. As the plotted result have a unique time axis with no time gap allowed, users can't observe time variation both in seconds and hours. If a procedure involves actions that must be achieved within the seconds and others within the hour, we suggest splitting the process in two according to the time scale.

Future Work

To improve the package, future work will focus on expanding ViSiElse's features. Increasing the number of groups or plotting nonlinear processes would help visualize more complex procedures involving parallel actions or interdisciplinary teams. Another improvement lies in the visualization of both quantitative and qualitative data. For example, qualitative data could be an indication of the goodness of the performance of the actions. Finally, the restriction on the number of individuals that can be plotted in a single graph for long actions could be solved by adding a gradient of color just like for the punctual actions.

In addition to the improvement of ViSiElse's features, future work will also aim at expanding the range of the allowed data type in ViSiElse by creating an option to change the x-axis unit. Users could visualize any quantitative variable. For example, concentrations or intensity could be visualized in different conditions. This improvement could extend the potential uses of ViSiElse. Similarly, developing an online interactive version of ViSiElse could facilitate its use and make it handy for any beginner 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 an instantaneous insight of individuals and/or groups' behavior. In a single one page graph, users check their raw data, visualize the time accuracy, compare groups and analyze the data distribution through statistical indicators and tests. ViSiElse's options are accessible through the ViSibook for all the actions characteristics (names, labels, types, order, delimitations and green and black zones), through the arguments of the "visielse" function for all the analysis features (time scale, groups, statistical indicators and tests) and through the arguments of the "plot" function for all the formatting options (labels size and color, adding a title, time interval size and unit). This graphical tool suits every process involving actions that can be linearized. It has been developed to be applied in the medical field but can be used more widely in all fields using time data to analyze behavior. With ViSiElse, data reliability and transparency can be assessed for an entire dataset in a single view.

Acknowledgements

The authors thank Pr Jean-Bernard Gouyon and Simon Lorrain for helpful comments and revising the manuscript. The University Hospital of La Réunion, for its support.

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

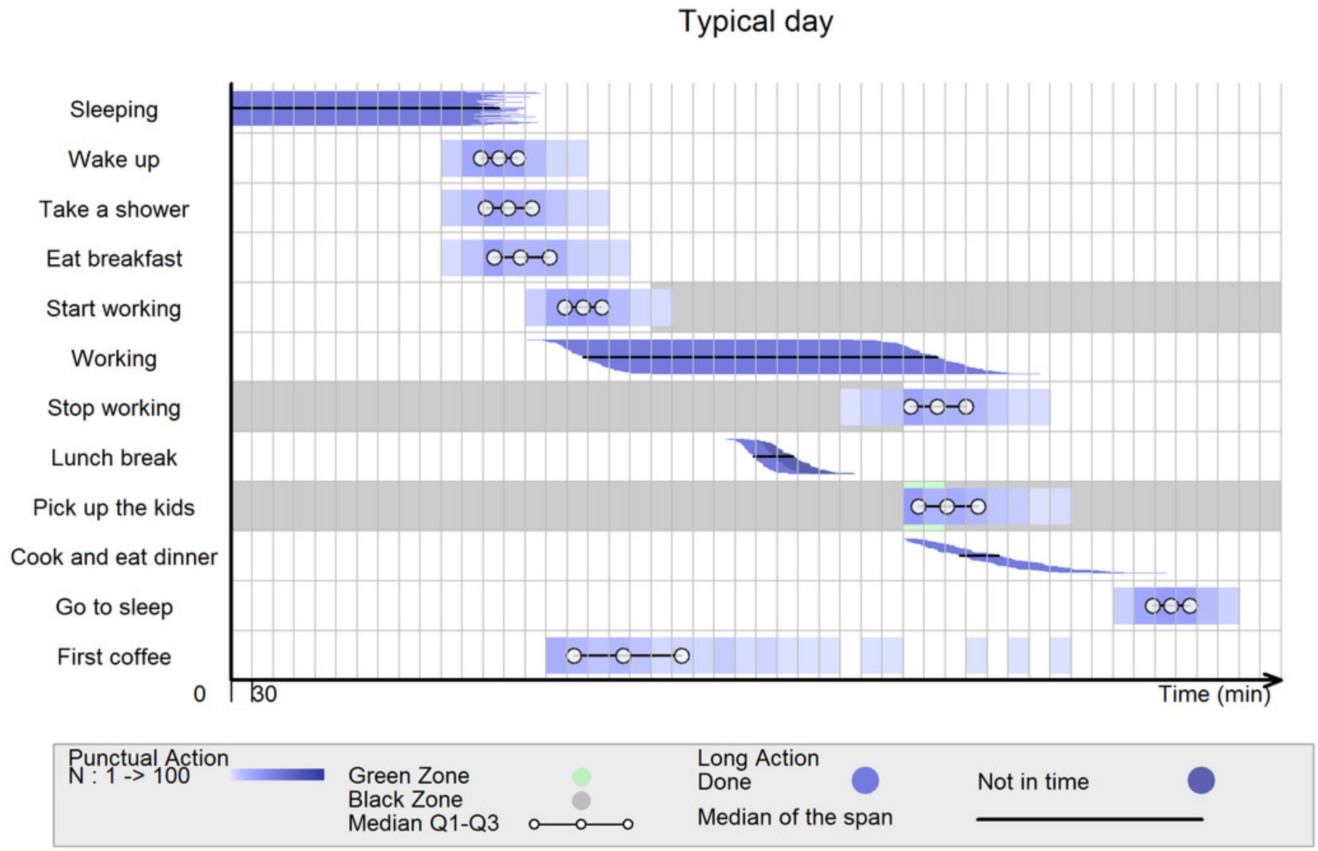
Figure 1

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 statistical indicators. 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. Statistical indicators 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 statistical indicators for long actions symbolized by a line proportional to the median duration of the actions.

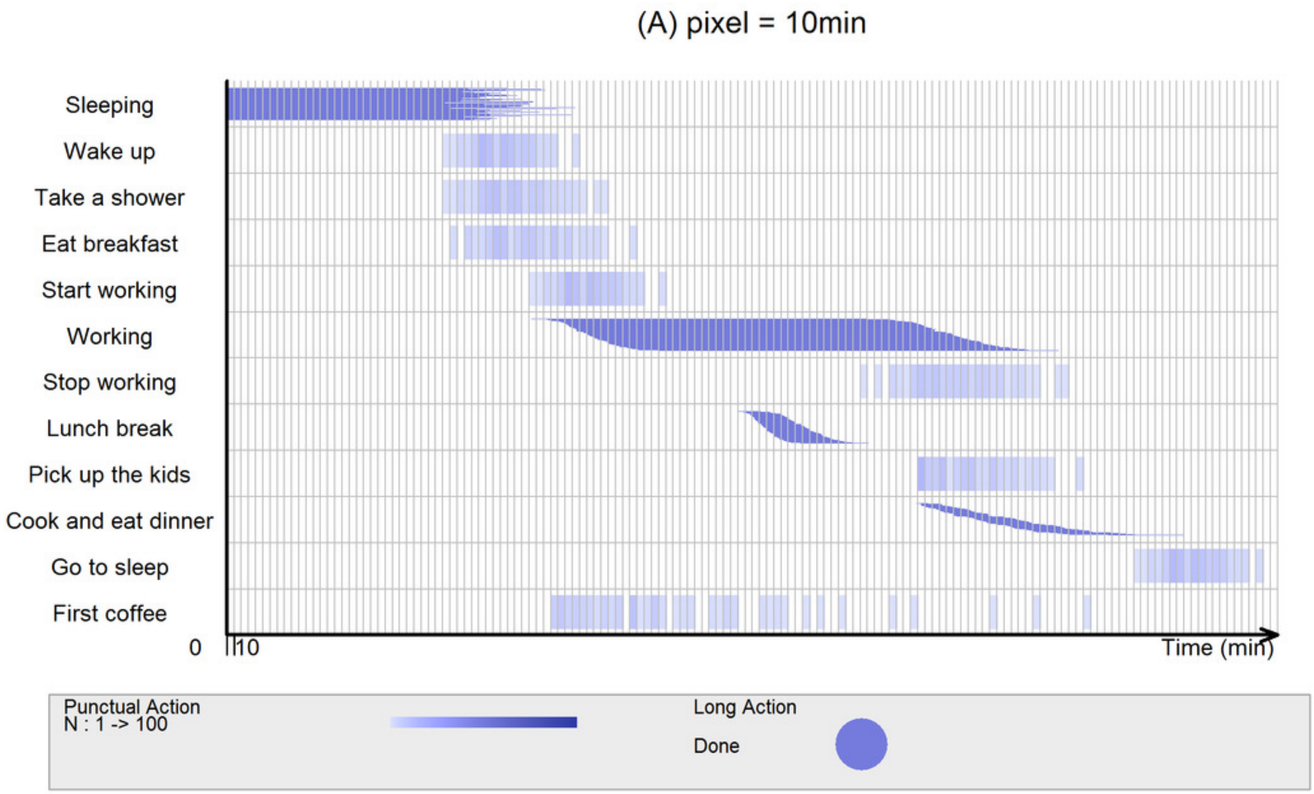


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 2019

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.


(A) pixel = 10min
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat dinner
Go to sleep
First coffee
0
10
Time (min)
Punctual Action
N : 1 -> 100
Long Action
Done

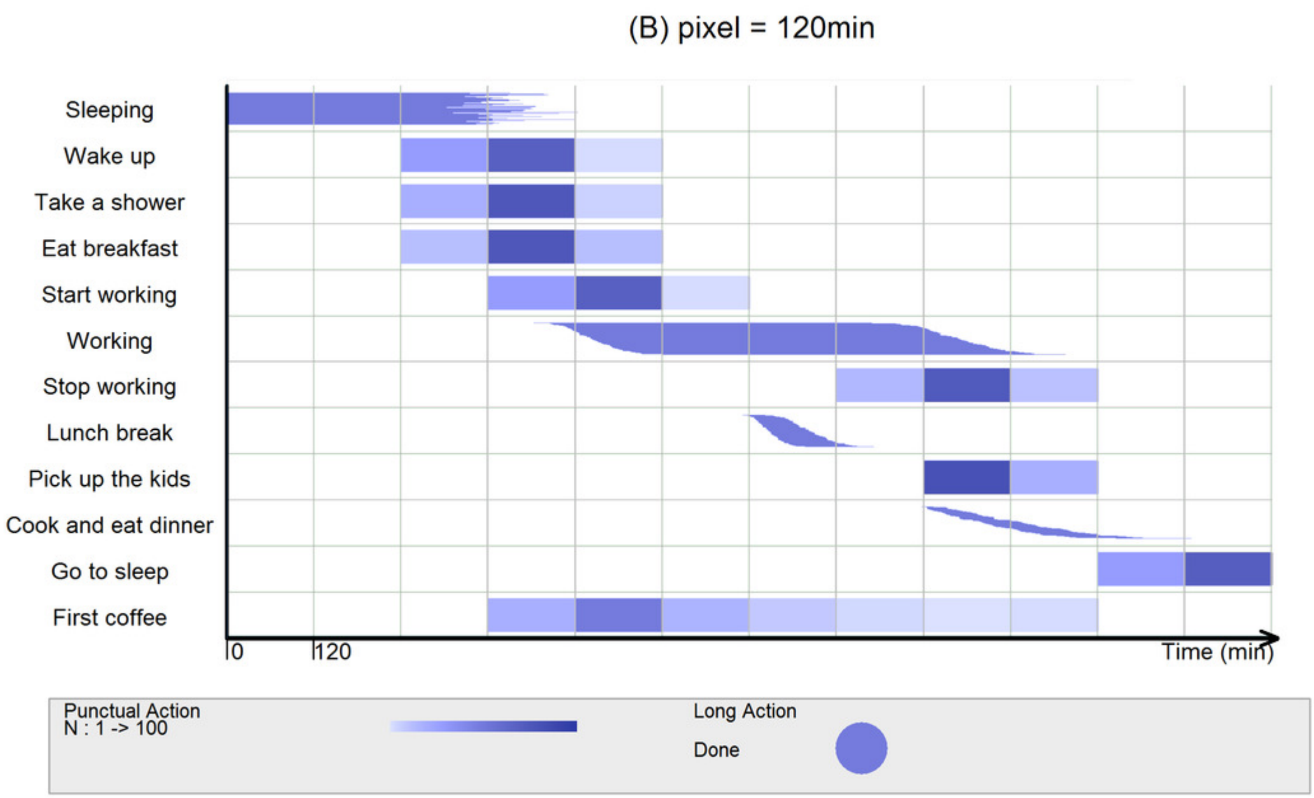

(B) pixel = 120min
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat dinner
Go to sleep
First coffee
0
120
Time (min)
Punctual Action
N : 1 -> 100
Long Action
Done

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.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 2019

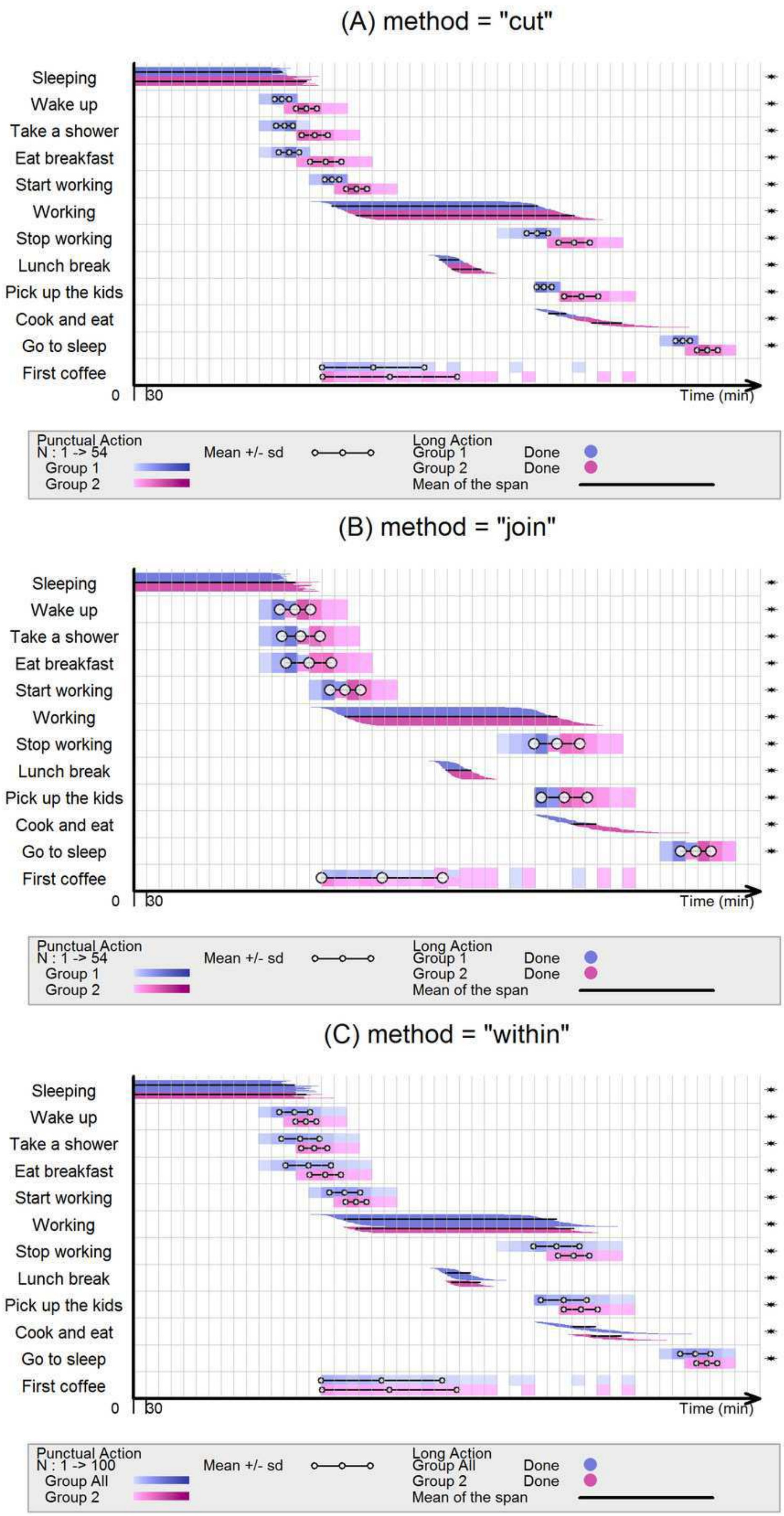

(A) method = "cut"
Sleeping
Wake up
Take a shower
Eat breakfast
Start working
Working
Stop working
Lunch break
Pick up the kids
Cook and eat
Go to sleep
First coffee
0
30
Time (min)
Punctual Action
N : 1 -> 54
Group 1
Group 2
Mean +/- sd
Long Action
Group 1 Done
Group 2 Done
Mean of the span
(B) method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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 2019

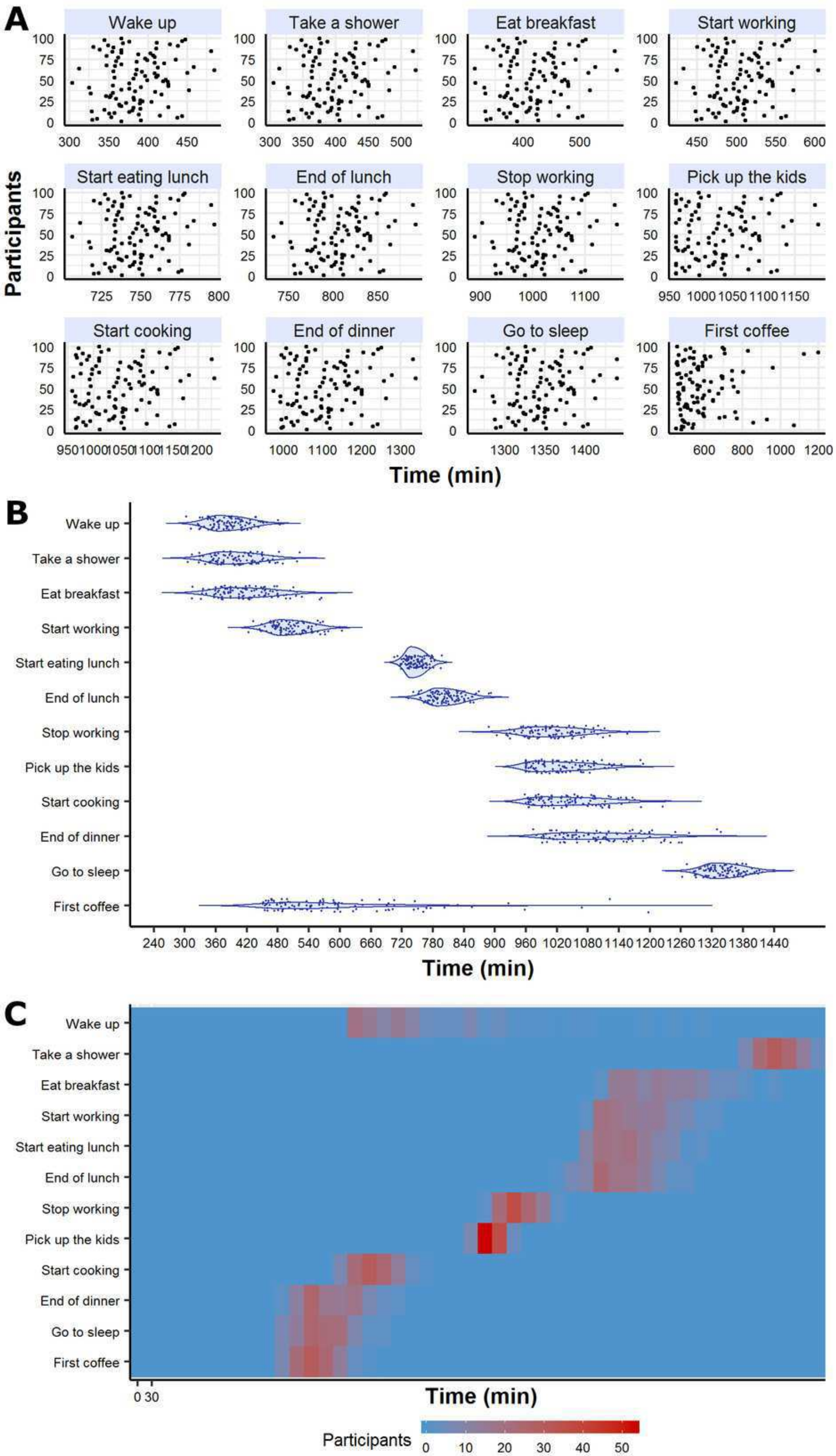

A
Wake up
Take a shower
Eat breakfast
Start working
Start eating lunch
End of lunch
Stop working
Pick up the kids
Start cooking
End of dinner
Go to sleep
First coffee
Participants
Time (min)
B
Time (min)
C
Time (min)
Participants
0
10
20
30
40
50

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").

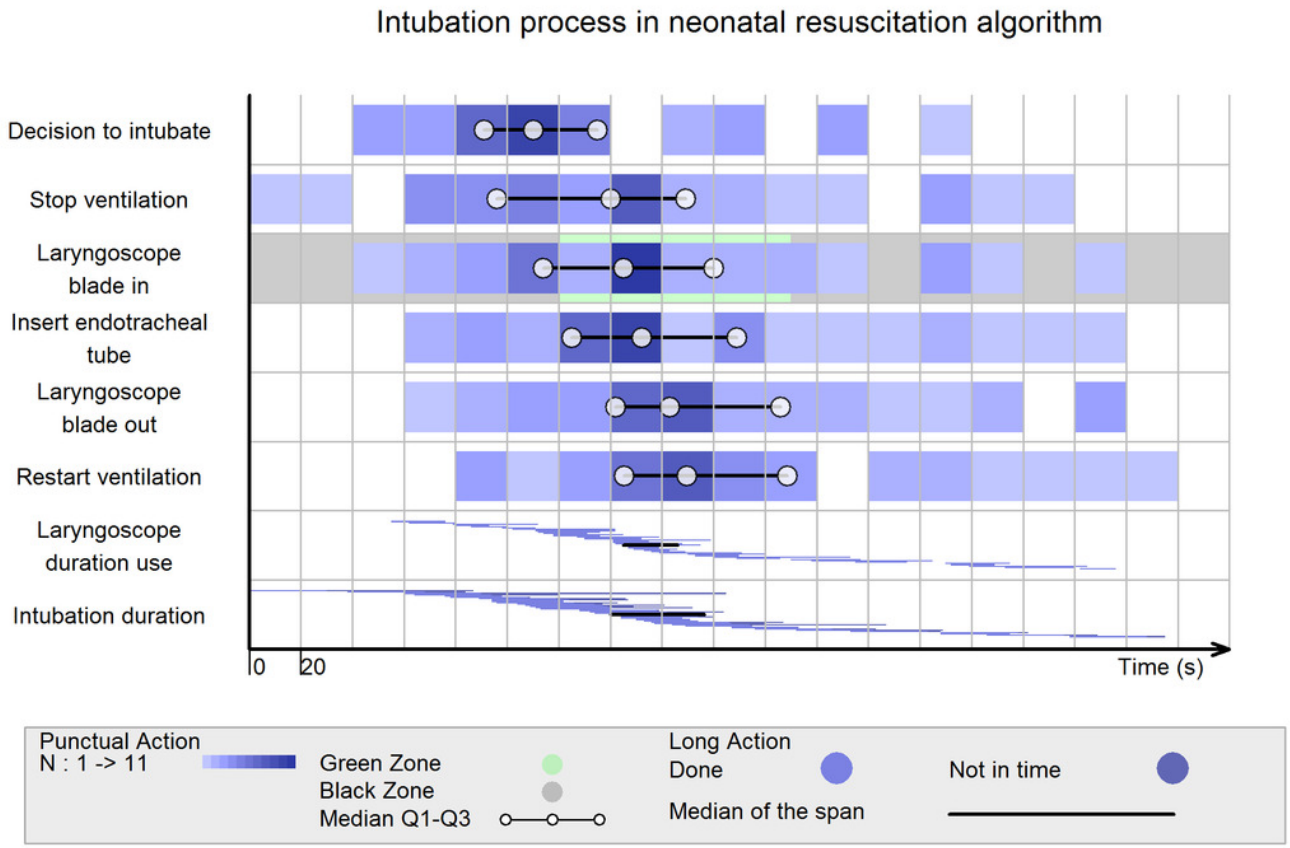


Figure 6

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).

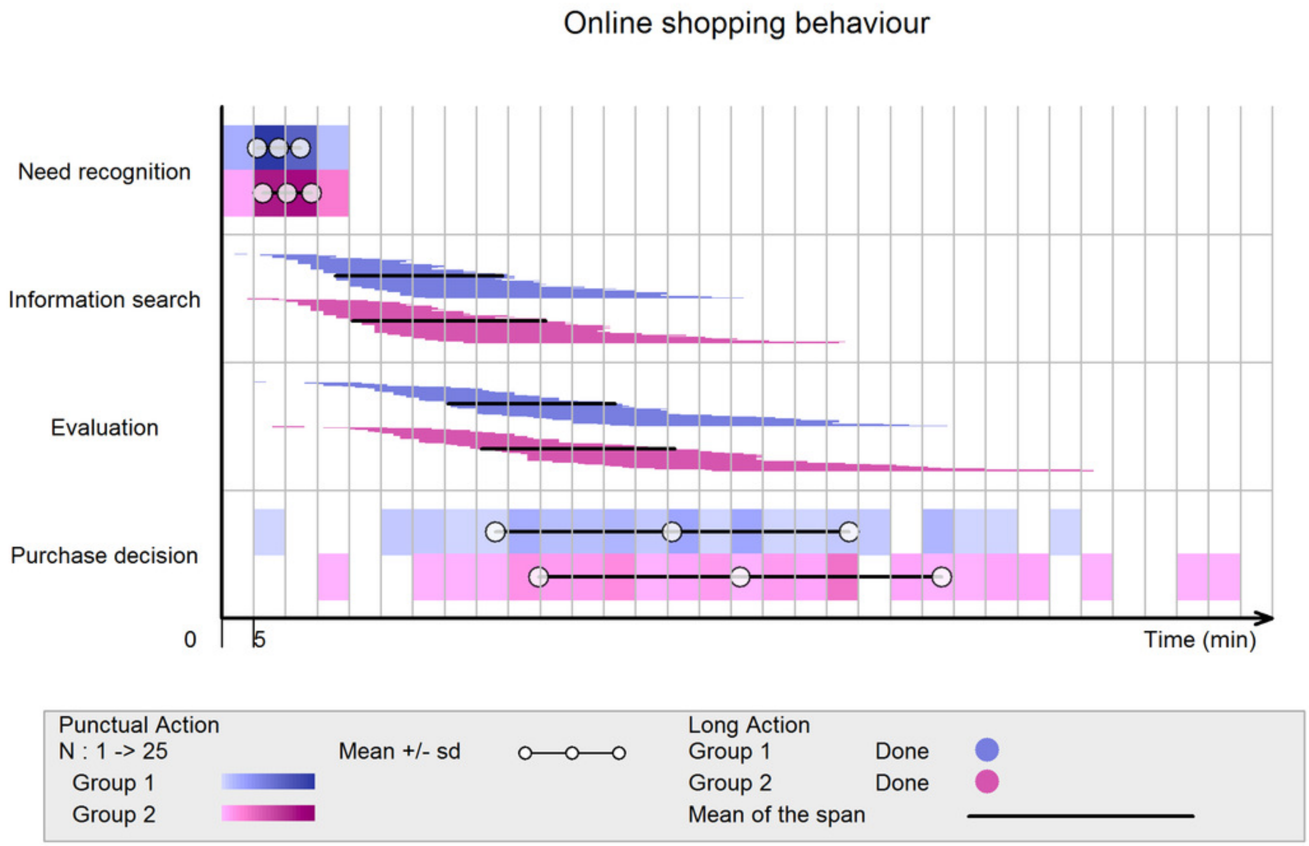


Table 1(on next page)

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.27665v2 | CC BY 4.0 Open Access | rec: 1 Aug 2019, publ: 1 Aug 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