# bupaR: Business Process Analysis in R

Gert Janssenswillen[1,2], Benoît Depaire[1]

UHasselt - Hasselt University, Faculty of Business Economics, Agoralaan, 3590
Diepenbeek, Belgium
Research Foundation Flanders (FWO), Egmontstraat 5, 1060 Brussels, Belgium
{gert.janssenswillen, benoit.depaire}@uhasselt.be

**Abstract.** This paper introduces `bupaR` which is a collection of R-
packages which can be used for exploring and visualizing event data and
monitoring processes. It creates a bridge between the BPM community
and the broader R and data science communities. Furthermore, it stimu-
lates the use of process mining in corporations, as R is a very wide used
analysis tool in business, with its popularity still strongly increasing.

**Keywords:** R, process analysis, process monitoring, data analysis,
statistics

## 1  Introduction

Over the past decades, the open source statistical language R [2] has seen an
enormous increase in popularity, not only among data science researchers, but
also within companies. One of the reasons for this rising popularity is the R-
package ecosystem on CRAN and github to which everyone can contribute. Re-
cently, the number of packages available on CRAN has exceeded 10.000. These
provide a huge range of functionalities, covering a diverse set of techniques and
applications.

This paper introduces bupaR, which is a set of R-packages for business pro-
cess analysis in R. The importance of these for the BPM community is twofold.

Firstly, it allows researchers and data analysts to tap into the wide range of
functionalities in the R-ecosystem. This does not only include traditional data
mining techniques, such as clustering, classification etc, but also new develop-
ments, such as text mining, spark, deep learning, etc. Furthermore, since R
originated as a statistical language, a vast amount of statistical tests becomes
available, allowing straightforward confirmatory analyses on event data.

Secondly, the introduced packages will allow to bridge the gap between busi-
ness and academics, as R is increasingly used as a tool for data analysis in cor-
porations. As such, the packages will provide a low threshold for companies to
discover the benefits of business process analysis and to become more acquainted
with the BPM community.

The next section will provide an overview of the packages and their func-
tionalities, while Section 3 discusses the maturity of the tools and some use
cases. Section 4 then provides more practical information on how to get started,
including a link to a screencast and to the website with further documentation.

## 2 Overview of functionalities

bupaR consists of different R-packages, each with their own purpose. The package bupaR itselfs acts as the central heart of the suite, providing the basic functionality for handling event data which is used by the other packages. The other packages are listed below, and are briefly introduced in the next paragraphs.

- **edeaR**: for exploratory and descriptive analysis of event data
- **xesreadR**: for reading and writing xes files
- **processmapR**: for creating process visualizations
- **processmonitoR**: for creating process dashboards for monitoring
- **eventdataR**: a data repository containing both real-life and artificial event logs

### 2.1 bupaR

The central package includes basic functionality for creating `eventlog`-objects in R. An event log in this context is a dataset combined with a *mapping*. The mapping contains the characteristics of the specific event log, i.e. the case identifier, activity identifier, timestamp, etc. Each element of this mapping refers to a variable in the data set. Figure 1 shows how an event log is created using the user interface of bupaR[1].
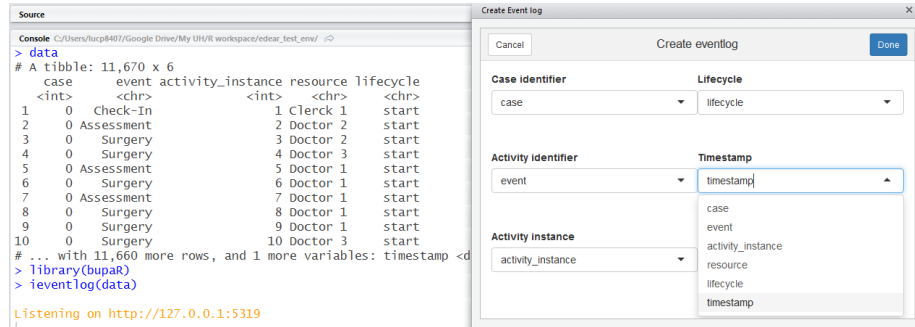


**Fig. 1.** Creating an event log object with bupaR

The package further contains several functions to get information about an event log. Furthermore, `bupaR` also provides specific event log versions for the generic R functions `print` and `summary`. For a complete list of all the functions, please look here: `http://bupar.net/bupar.html`.

---

[1] The function used in Figure 1 is called `ieventlog`, in which the first letter `i` indicates that it opens a user interface, which is useful when doing interactive analyis. However, each of these interface functions has a non-interface alternative (in this case the function `eventlog`, which can be used both interactively and for scripting).

## 2.2 edeaR

The name of the package `edeaR` stands for Exploratory and Descriptive Event-data Analysis in R. It was first introduced at SIMPDA 2016 [4] and is the oldest member of the `bupaR` family, although it has gone through important improvements and extensions recently. As itsname implies, the purpose of this package is to perform more in-depth analyses of event logs. It basically contains two sets of functions: *metrics* for the analysis of data and *filters* for data subsetting.

The metric functions are largely build around the concepts of Lean Six Sigma en Operational Excellence [6], i.e. metrics concerning time (processing time, throughput time, idle time), variance (trace length, trace coverage, start activities, etc.), rework (repetitions of work, selfloops of activites), and finally also concerning organizational aspects (e.g. involvement and specialization of resources) [7]. Furthermore, each of the available metrics can be computed at different levels of granularity: e.g. for the complete log, for specific cases, for specific resources, etc.

While each of the metrics returns numeric results (an output table or a set of numbers), the generic `plot` function of R has been customized to provided tailored plots for each of the metrics at each level of granularity [8].

The subsetting functions can be found easily as each of them starts with the prefix `filter_`. Different filter functions are available: filtering specific case identifiers, specific attributes (event or case attributes), filtering on throughput time, filtering activities based on their frequencies, etc. Note that for each of the filter functions, also an interface is provided for straightforward interactive use. A complete list of functions and guide for `edeaR` can be found here: `http://bupar.net/edeaR.html`.

## 2.3 processmapR

The package `processmapR` provides a straightforward way to create visualizations of processes, using customizable *map profiles*. The default profile is the frequency profile with absolute frequencies on both arcs and nodes, as can be seen in the example in Figure 2. The return object is a `dgr_graph` object from the `DiagrammeR` package, and can be further customized by the user if needed. Complex graphs can be simplified by combining the processmapR functions with subsetting functions of `edeaR`. More information can be found `http://bupar.net/processmapR.html`.

## 2.4 xesreadR

As is clear from the name, the purpose of `xesreadR` is to read and write xes-files. This makes `bupar` compatible with the IEEE standard for sharing and storing event data, and thus with other process mining tools. For example, this is useful if you want to combine `bupaR` with RapidProM, as RapidMiner also supports the use of R-scripts.
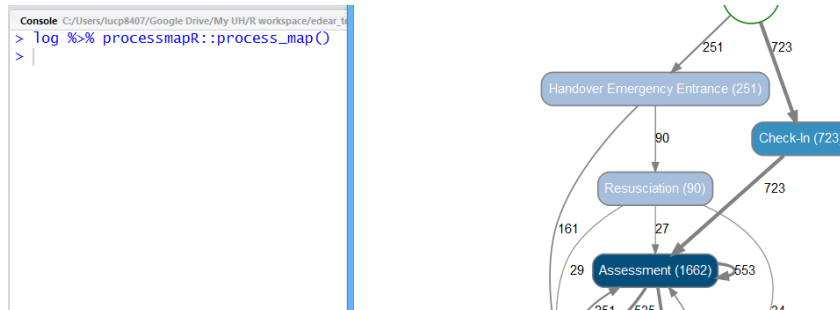
**Fig. 2.** Creating a process map with processmapR

### 2.5 eventdataR

With the purpose of testing new techniques and algorithms, the package `eventdataR` contains a data repository of both artificial and real-life event log. Each of them can be loaded very straightforwardly with the `data` function in R. The available datasets are listed here: `http://bupar.net/eventdataR.html`.

### 2.6 processmonitoR

While the other packages can be used both interactively and in production, `processmonitoR` is mostly intended for the latter case, as it provides building blocks for online process monitoring dashboards. The available dashboards, which build upon the Shiny framework [1], are largely organized along the metrics of `edeaR`.

## 3 Maturity and Use Cases

As explained, the packages can be used for several purposes, ranging from a one-time interactive analysis of event data to building a full-fledged online process monitoring dashboard. `edeaR`, the oldest package included in `bupaR` was first published on CRAN in January 2016 and has since been downloaded more than 5000 times wordwide. Its use is also illustrated in several submissions for the BPI challenge [3,5].

Although the other packages have been shared publicly more recently, the underlying functionality has already proved useful in several projects. A process monitoring dashboard was developed to analyze the process of connecting new customers at a Belgian utility network operator, while the visualizations of processes have been used for visualizing train deviations at a European railway infrastructure manager.

# 4   Getting started

**Screencast** A screencast discussing the installation and use of `bupaR` and the related packages can be found here: `https://goo.gl/huHTGE`.

**Prerequisites** In order to use `bupaR`, R and Rstudio (or another IDE) need to be installed. More information can be found on `cran.rstudio.com` and `www.rstudio.com`.

**Installing bupaR** Installing `bupaR` can be done easily by executing `install.packages("bupaR")` in the R-console. You can then load it using `library(bupaR)`. The first time you will be asked to also install the related packages, on which you should answer Yes (Y).

**Further guidance** The website `www.bupar.net` contains ample documentation and examples of the tools discussed.

## Acknowledgements

Special thanks and appreciation go to Mieke Jans and Marijke Swennen for catalyzing the development of the initial version of edeaR and for their suggestions and feedback during the further evolution of bupaR.

## References

1. Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J.: shiny: Web application framework for r, 2015. URL http://CRAN. R-project. org/package= shiny. R package version 0.11 (2015)
2. Ihaka, R., Gentleman, R.: R: a language for data analysis and graphics. Journal of computational and graphical statistics 5(3), 299–314 (1996)
3. Janssenswillen, G., Creemers, M., Jouck, T., Martin, N., Swennen, M.: Does werk. nl work? (2016)
4. Janssenswillen, G., Swennen, M., Depaire, B., Jans, M., Vanhoof, K.: Enabling event-data analysis in r: Demonstration. RWTH Aachen University (2015)
5. Martin, N., Janssenswillen, G., Jouck, T., Swennen, M., Hosseinpour, M., Masoumigoudarzi, F.: An exploration and analysis of the building permit application process in five dutch municipalities (2015)
6. Swennen, M., Janssenswillen, G., Jans, M., Depaire, B., Vanhoof, K.: Capturing process behavior with log-based process metrics. In: SIMPDA. pp. 141–144 (2015)
7. Swennen, M., Martin, N., Janssenswillen, G., Jans, M.J., Depaire, B., Caris, A., Vanhoof, K.: Capturing resource behaviour from event logs. RWTH Aachen University (2016)
8. Wickham, H.: ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3(2), 180–185 (2011)