Check for updates

SOFTWARE TOOL ARTICLE

REVISED epicontacts: Handling, visualisation and analysis of epidemiological contacts [version 2; peer review: 2 approved, 1 approved with reservations]

VP Nagraj [1], Nistara Randhawa[2], Finlay Campbell [3], Thomas Crellen [4], Bertrand Sudre[5], Thibaut Jombart [3]

[1]Research Computing, University of Virginia School of Medicine, Charlottesville, VA, 22903, USA
[2]One Health Institute, University of California, Davis, Davis, CA, 95616, USA
[3]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, W2 1PG, UK
[4]Mahidol-Oxford Tropical Medicine Research Unit, Bangkok , 10400, Thailand
[5]European Centre for Disease Prevention and Control, Stockholm, Sweden

Abstract
Epidemiological outbreak data is often captured in line list and contact format to facilitate contact tracing for outbreak control. *epicontacts* is an R package that provides a unique data structure for combining these data into a single object in order to facilitate more efficient visualisation and analysis. The package incorporates interactive visualisation functionality as well as network analysis techniques. Originally developed as part of the Hackout3 event, it is now developed, maintained and featured as part of the R Epidemics Consortium (RECON). The package is available for download from the Comprehensive R Archive Network (CRAN) and GitHub.

Keywords
contact tracing, outbreaks, R

This article is included in the RPackage gateway.

This article is included in the R Epidemics Consortium (RECON) collection.

Open Peer Review

Reviewer Status ? ✔ ✔

	Invited Reviewers		
	1	**2**	**3**
REVISED **version 2** published 11 Oct 2018			✔ report
version 1 published 10 May 2018	? report	✔ report	

1 **Melissa A. Rolfes**, Centers for Disease Control and Prevention (CDC) , Atlanta, USA

2 **Peter Adebayo Adewuyi**, Liberia Field Epidemiology Training Program, Monrovia, Liberia
African Field Epidemiology Network (AFENET), Kampala, Uganda

3 **Josie Athens** , University of Otago, Dunedin, New Zealand

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: VP Nagraj (vpnagraj@virginia.edu)

Author roles: Nagraj V: Conceptualization, Software, Writing – Original Draft Preparation; **Randhawa N**: Conceptualization, Software, Writing – Original Draft Preparation; **Campbell F**: Conceptualization, Software, Writing – Original Draft Preparation; **Crellen T**: Conceptualization, Software; **Sudre B**: Conceptualization; **Jombart T**: Conceptualization, Software, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

> **REVISED** **Amendments from Version 1**
>
> In response to suggestions provided during the peer review process, the authors have made several updates to the manuscript. The body of the text is now organized more intuitively, introducing use cases for the *epicontacts* package before discussing specific functionality. Furthermore, the provenance of the data set is now described in the "Data handling" sub-section. The text has also been updated to include links to additional resources that demonstrate package usage. The authors feel that these changes have improved the manuscript and would like to thank the reviewers for providing their feedback.
>
> **See referee reports**

Introduction

In order to study, prepare for, and intervene against disease outbreaks, infectious disease modellers and public health professionals need an extensive data analysis toolbox. Disease outbreak analytics involve a wide range of tasks that need to be linked together, from data collection and curation to exploratory analyses, and more advanced modelling techniques used for incidence forecasting[1,2] or to predict the impact of specific interventions[3,4]. Recent outbreak responses suggest that for such analyses to be as informative as possible, they need to rely on a wealth of available data, including timing of symptoms, characterisation of key delay distributions (e.g. incubation period, serial interval), and data on contacts between patients[5–8].

The latter type of data is particularly important for outbreak analysis, not only because contacts between patients are useful for unravelling the drivers of an epidemic[9,10], but also because identifying new cases early can reduce ongoing transmission via contact tracing, i.e. follow-up of individuals who reported contacts with known cases[11,12]. However, curating contact data and linking them to existing line lists of cases is often challenging, and tools for storing, handling, and visualising contact data are often missing[13,14].

Here, we introduce `epicontacts`, an R[15] package providing a suite of tools aimed at merging line lists and contact data, and providing basic functionality for handling, visualising and analysing epidemiological contact data. Maintained as part of the R Epidemics Consortium (RECON), the package is integrated into an ecosystem of tools for outbreak response using the R language.

Use cases

Those interested in using `epicontacts` should have a line list of cases as well as a record of contacts between individuals. Both datasets must be enumerated in tabular format with rows and columns. At minimum the line list requires one column with a unique identifier for every case. The contact list needs two columns for the source and destination of each pair of contacts. The datasets can include arbitrary features of case or contact beyond these columns. Once loaded into R and stored as `data.frame` objects, these datasets can be passed to the `make_epicontacts()` function (see 'Methods' section for more detail). For an example of data prepared in this format, users can refer to the `outbreaks` R package.

```
# load the outbreaks package
library(outbreaks)

# example simulated ebola data

# line list
str(ebola_sim$linelist)

## 'data.frame':    5888 obs. of 9 variables:
## $ case_id             : chr  "d1fafd" "53371b" "f5c3d8" "6c286a" ...
## $ generation          : int  0 1 1 2 2 0 3 3 2 3 ...
## $ date_of_infection   : Date, format: NA "2014-04-09" ...
## $ date_of_onset       : Date, format: "2014-04-07" "2014-04-15" ...
## $ date_of_hospitalisation: Date, format: "2014-04-17" "2014-04-20" ...
## $ date_of_outcome     : Date, format: "2014-04-19" NA ...
## $ outcome             : Factor w/ 2 levels "Death","Recover": NA NA 2 1 2 NA 2 1 2 1 ...
## $ gender              : Factor w/ 2 levels "f","m": 1 2 1 1 1 1 1 1 2 2 ...
## $ hospital            : Factor w/ 11 levels "Connaught Hopital",..: 4 2 7 NA 7 NA 2 9 7 11 ...

# contact list
str(ebola_sim$contacts)

## 'data.frame':    3800 obs. of  3 variables:
## $ infector: chr  "d1fafd" "cac51e" "f5c3d8" "0f58c4" ...
```

```
## $ case_id : chr  "53371b" "f5c3d8" "0f58c4" "881bd4" ...
## $ source  : Factor w/ 2 levels "funeral","other": 2 1 2 2 2 1 2 2 2 2 ...
```

```
# example middle east respiratory syndrome data

# line list
str(mers_korea_2015$linelist)
```

```
## 'data.frame':    162 obs. of 15 variables:
## $ id            : chr  "SK_1" "SK_2" "SK_3" "SK_4" ...
## $ age           : int  68 63 76 46 50 71 28 46 56 44 ...
## $ age_class     : chr  "60-69" "60-69" "70-79" "40-49" ...
## $ sex           : Factor w/ 2 levels "F","M": 2 1 2 1 2 2 1 1 2 2 ...
## $ place_infect  : Factor w/ 2 levels "Middle East",..: 1 2 2 2 2 2 2 2 2 2 ...
## $ reporting_ctry: Factor w/ 2 levels "China","South Korea": 2 2 2 2 2 2 2 2 2 1 ...
## $ loc_hosp      : Factor w/ 13 levels "365 Yeollin Clinic, Seoul",..: 10 10 10 10 1 10 10 13 10 10 ...
## $ dt_onset      : Date, format: "2015-05-11" "2015-05-18" ...
## $ dt_report     : Date, format: "2015-05-19" "2015-05-20" ...
## $ week_report   : Factor w/ 5 levels "2015_21","2015_22",..: 1 1 1 2 2 2 2 2 2 2 ...
## $ dt_start_exp  : Date, format: "2015-04-18" "2015-05-15" ...
## $ dt_end_exp    : Date, format: "2015-05-04" "2015-05-20" ...
## $ dt_diag       : Date, format: "2015-05-20" "2015-05-20" ...
## $ outcome       : Factor w/ 2 levels "Alive","Dead": 1 1 2 1 1 2 1 1 1 1 ...
## $ dt_death      : Date, format: NA NA ...
```

```
# contact list
str(mers_korea_2015$contacts)
```

```
## 'data.frame':    98 obs. of  4 variables:
## $ from         : chr  "SK_14" "SK_14" "SK_14" "SK_14" ...
## $ to           : chr  "SK_113" "SK_116" "SK_41" "SK_112" ...
## $ exposure     : Factor w/ 5 levels "Contact with HCW",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ diff_dt_onset: int  10 13 14 14 15 15 15 16 16 16 ...
```

The data handling, visualization, and analysis methods described above represent the bulk of epicontacts features. More examples of how the package can be used as well as demonstrations of additional features can be found through the RECON learn platform and the epicontacts vignettes.

Methods
Operation
epicontacts is released as an open-source R package. A stable release is available for Windows, Mac and Linux operating systems via the CRAN repository. The latest development version of the package is available through the RECON Github organization. At minimum users must have R installed. No other system dependencies are required.

```
# install from CRAN
install.packages("epicontacts")

# install from Github
install.packages("devtools")
devtools::install_github("reconhub/epicontacts")

# load and attach the package
library(epicontacts)
```

Implementation
Data handling. epicontacts includes a novel data structure to accommodate line list and contact list datasets in a single object. This object is constructed with the make_epiconctacts() function and includes attributes from the original datasets. Once combined, these are mapped internally in a graph paradigm as nodes and edges. The epicontacts data structure also includes a logical attribute for whether or not this resulting network is directed.

The package takes advantage of R's generic functions, which call specific methods depending on the class of an object. This is implemented several places, including the summary.epicontacts() and print.epicontacts() methods, both of which are respectively called when the summary() or print() functions are used on an

epicontacts object. The package does not include built-in data, as exemplary contact and line list datasets are available in the outbreaks package[16].

The example that follows will use the mers_korea_2015 dataset from outbreaks, which which includes initial data collected by the Epidemic Intelligence group at European Centre for Disease Prevention and Control (ECDC) during the 2015 outbreak of Middle East respiratory syndrome (MERS-CoV) in South Korea. Note that the data used here was provided in outbreaks for teaching purposes, and therefore does not include the complete line list or contacts from the outbreak.

```
# install the outbreaks package for data
install.packages("outbreaks")

# load the outbreaks package
library(outbreaks)

# construct an epicontacts object
x <- make_epicontacts(linelist=mers_korea_2015[[1]],
                      contacts = mers_korea_2015[[2]],
                      directed=TRUE)
# print the object
x


##
## /// Epidemiological Contacts ///
##
## // class: epicontacts
## // 162 cases in linelist; 98 contacts;  directed
##
## // linelist
##
## 'data.frame':    162 obs. of 15 variables:
## $ id           : chr "SK_1" "SK_2" "SK_3" "SK_4" ...
## $ age          : int 68 63 76 46 50 71 28 46 56 44 ...
## $ age_class    : chr "60-69" "60-69" "70-79" "40-49" ...
## $ sex          : Factor w/ 2 levels "F","M": 2 1 2 1 2 2 1 1 2 2 ...
## $ place_infect : Factor w/ 2 levels "Middle East",..: 1 2 2 2 2 2 2 2 2 2 ...
## $ reporting_ctry: Factor w/ 2 levels "China","South Korea": 2 2 2 2 2 2 2 2 2 1 ...
## $ loc_hosp     : Factor w/ 13 levels "365 Yeollin Clinic, Seoul",..: 10 10 10 10 1 10 10 13 10 10 ...
## $ dt_onset     : Date, format: "2015-05-11" "2015-05-18" ...
## $ dt_report    : Date, format: "2015-05-19" "2015-05-20" ...
## $ week_report  : Factor w/ 5 levels "2015_21","2015_22",..: 1 1 1 2 2 2 2 2 2 2 ...
## $ dt_start_exp : Date, format: "2015-04-18" "2015-05-15" ...
## $ dt_end_exp   : Date, format: "2015-05-04" "2015-05-20" ...
## $ dt_diag      : Date, format: "2015-05-20" "2015-05-20" ...
## $ outcome      : Factor w/ 2 levels "Alive","Dead": 1 1 2 1 1 2 1 1 1 1 ...
## $ dt_death     : Date, format: NA NA ...
##
## // contacts
##
## 'data.frame':    98 obs. of  4 variables:
## $ from         : chr  "SK_14" "SK_14" "SK_14" "SK_14" ...
## $ to           : chr  "SK_113" "SK_116" "SK_41" "SK_112" ...
## $ exposure     : Factor w/ 5 levels "Contact with HCW",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ diff_dt_onset: int  10 13 14 14 15 15 15 16 16 16 ...

# view a summary of the object
summary(x)


##
## /// Overview //
##   // number of unique IDs in linelist: 162
##   // number of unique IDs in contacts: 97
##   // number of unique IDs in both: 97
##   // number of contacts: 98
##   // contacts with both cases in linelist: 100 %
```

```
##
## /// Degrees of the network //
##    // in-degree summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    1.00    1.00    1.01    1.00    3.00
##
##    // out-degree summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    1.01    0.00   38.00
##
##    // in and out degree summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   1.000   1.000   2.021   1.000  39.000
##
## /// Attributes //
##    // attributes in linelist:
##  age age_class sex place_infect reporting_ctry loc_hosp dt_onset dt_report week_report
dt_start_exp dt_end_exp dt_diag outcome dt_death
##
##    // attributes in contacts:
##  exposure diff_dt_onset
```

Data visualisation. `epicontacts` implements two interactive network visualisation packages: `visNetwork` and `threejs`[17,18]. These frameworks provide R interfaces to the `vis.js` and `three.js` JavaScript libraries respectively. Their functionality is incorporated in the generic `plot()` method (Figure 1) for an `epicontacts` object, which can be toggled between either with the "type" parameter. Alternatively, the `visNetwork` interactivity is accessible via `vis_epicontacts()` (Figure 2), and `threejs` through `graph3D()`(Figure 3). Each function has a series of arguments that can also be passed through `plot()`. Both share a color palette, and users can specify node, edge and background colors. However, `vis_epicontacts()`
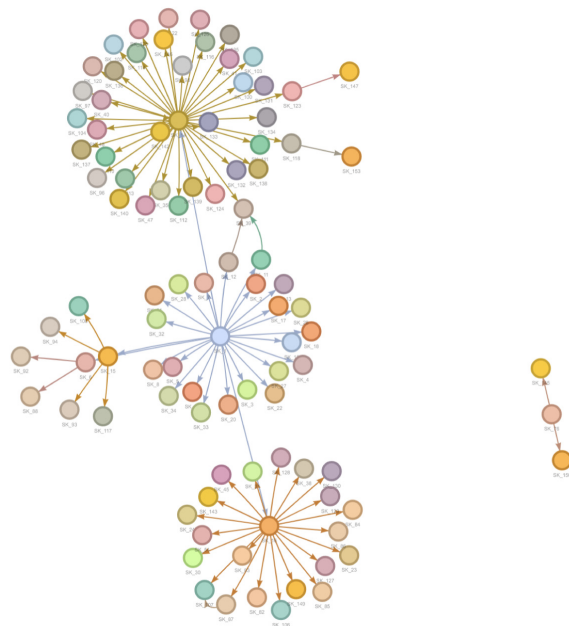

Figure 1. The generic plot() method for an epicontacts object will use the visNetwork method by default.
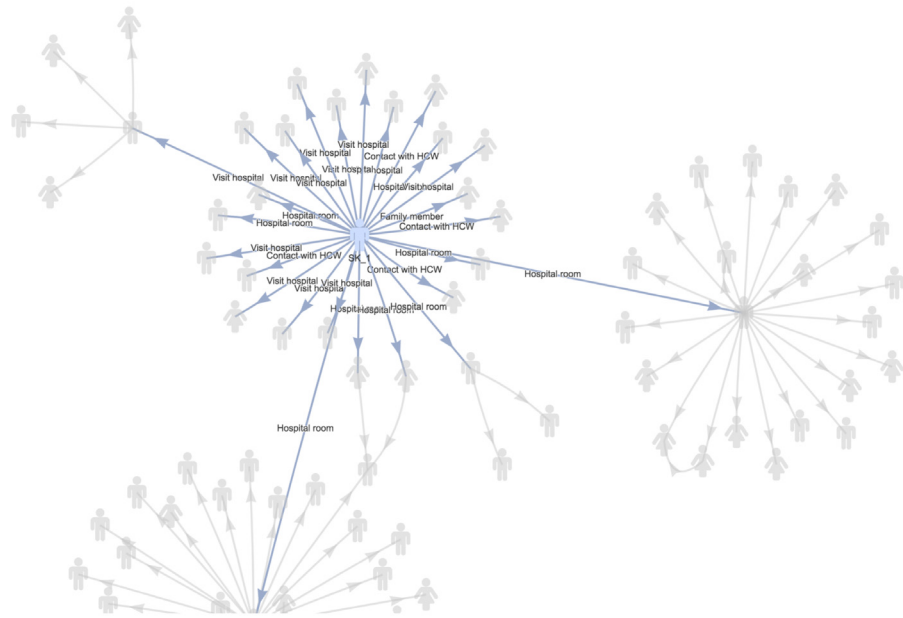
SK_1

Figure 2. The vis_epicontacts() function explicitly calls visNetwork to make an interactive plot of the contact network.
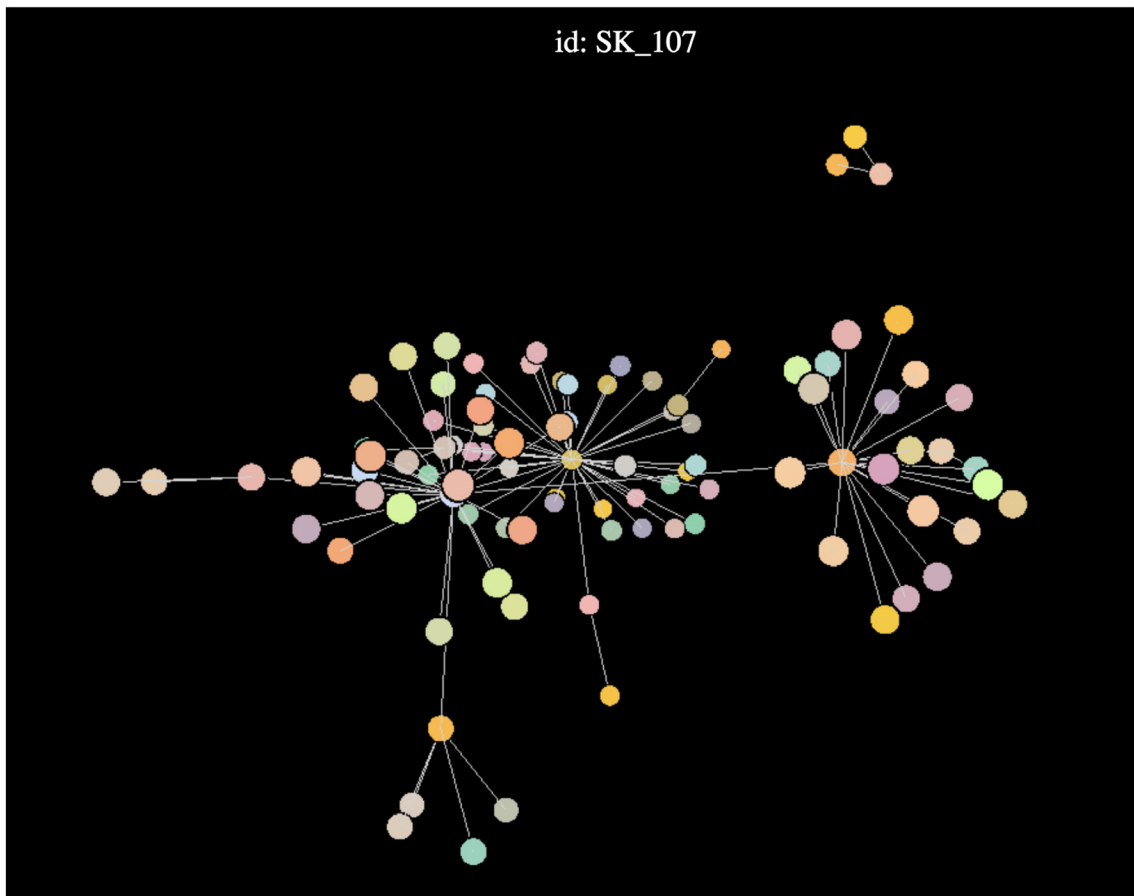

Figure 3. The graph3D() function generates a three-dimensional network plot.

includes a specification for "node_shape" by a line list attribute as well as a customization of that shape with an icon from the Font Awesome icon library. The principal distinction between the two is that `graph3D()` is a three-dimensional visualisation, allowing users to rotate clusters of nodes to better inspect their relationships.

```
plot(x)

vis_epicontacts(x,
                node_shape = "sex",
                shapes = c(F = "female", M = "male"),
                edge_label = "exposure")


graph3D(x, bg_col = "black")
```

Data analysis. Subsetting is a typical preliminary step in data analysis. `epicontacts` leverages a customized `subset` method to filter line lists or contacts based on values of particular attributes from nodes, edges or both. If users are interested in returning only contacts that appear in the line list (or vice versa), the `thin()` function implements such logic.

```
# subset for males
subset(x, node_attribute = list("sex" = "M"))

# subset for exposure in emergency room
subset(x, edge_attribute = list("exposure" = "Emergency room"))

# subset for males who survived and were exposed in emergency room
subset(x,
       node_attribute = list("sex" = "M", "outcome" = "Alive"),
       edge_attribute = list("exposure" = "Emergency room"))

thin(x, "contacts")
thin(x, "linelist")
```

For analysis of pairwise contact between individuals, the `get_pairwise()` feature searches the line list based on the specified attribute. If the given column is a numeric or date object, the function will return a vector containing the difference of the values of the corresponding "from" and "to" contacts. This can be particularly useful, for example, if the line list includes the date of onset of each case. The subtracted value of the contacts would approximate the serial interval for the outbreak[19]. For factors, character vectors and other non-numeric attributes, the default behavior is to print the associated line list attribute for each pair of contacts. The function includes a further parameter to pass an arbitrary function to process the specified attributes. In the case of a character vector, this can be helpful for tabulating information about different contact pairings with `table()`.

```
# find interval between date onset in cases
get_pairwise(x, "dt_onset")

# find pairs of age category contacts
get_pairwise(x, "age_class")

# tabulate the pairs of age category contacts
get_pairwise(x, "age_class", f = table)
```

Discussion
Benefits
While there are software packages available for epidemiological contact visualisation and analysis, none aim to accommodate line list and contact data as purposively as `epicontacts`[20–22]. Furthermore, this package strives to solve a problem of plotting dense graphs by implementing interactive network visualisation tools. A static plot of a network with many nodes and edges may be difficult to interpret. However, by rotating or hovering over an `epicontacts` visualisation, a user may better understand the data.

Future considerations
The maintainers of `epicontacts` anticipate new features and functionality. Future development could involve performance optimization for visualising large networks, as generating these interactive plots is resource intensive. Additionally, attention may be directed towards inclusion of alternative visualisation methods.

Conclusions
`epicontacts` provides a unified interface for processing, visualising and analyzing disease outbreak data in the R language. The package and its source are freely available on CRAN and GitHub. By developing functionality

with line list and contact list data in mind, the authors aim to enable more efficient epidemiological outbreak analyses.

Software availability

Software available from: https://CRAN.R-project.org/package=epicontacts

Source code available from: https://github.com/reconhub/epicontacts

Archived source code as at time of publication: https://zenodo.org/record/1210993[23]

Software license: GPL 2

References

1. Funk S, Camacho A, Kucharski AJ, *et al.*: **Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model.** *Epidemics.* 2018; **22**: 56–61.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Nouvellet P, Cori A, Garske T, *et al.*: **A simple approach to measure transmissibility and forecast incidence.** *Epidemics.* 2018; **22**: 29–35.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Nouvellet P, Garske T, Mills HL, *et al.*: **The role of rapid diagnostics in managing Ebola epidemics.** *Nature.* 2015; **528**(7580): S109–116.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Parker EP, Molodecky NA, Pons-Salort M, *et al.*: **Impact of inactivated poliovirus vaccine on mucosal immunity: implications for the polio eradication endgame.** *Expert Rev Vaccines.* 2015; **14**(8): 1113–1123.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Cauchemez S, Fraser C, Van Kerkhove MD, *et al.*: **Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility.** *Lancet Infect Dis.* 2014; **14**(1): 50–56.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. WHO Ebola Response Team, Aylward B, Barboza P, *et al.*: **Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections.** *N Engl J Med.* 2014; **371**(16): 1481–1495.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. WHO Ebola Response Team, Agua-Agum J, Ariyarajah A, *et al.*: **West African Ebola epidemic after one year--slowing but not yet under control.** *N Engl J Med.* 2015; **372**(6): 584–587.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Cori A, Donnelly CA, Dorigatti I, *et al.*: **Key data for outbreak evaluation: building on the Ebola experience.** *Philos Trans R Soc Lond B Biol Sci.* 2017; **372**(1721): pii: 20160371.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. International Ebola Response Team, Agua-Agum J, Ariyarajah A, *et al.*: **Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study.** *PLoS Med.* 2016; **13**(11): e1002170.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Cauchemez S, Nouvellet P, Cori A, *et al.*: **Unraveling the drivers of MERS-CoV transmission.** *Proc Natl Acad Sci U S A.* 2016; **113**(32): 9081–9086.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Senga M, Koi A, Moses L, *et al.*: **Contact tracing performance during the Ebola virus disease outbreak in Kenema district, Sierra Leone.** *Philos Trans R Soc Lond B Biol Sci.* 2017; **372**(1721): pii: 20160300.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Saurabh S, Prateek S: **Role of contact tracing in containing the 2014 Ebola outbreak: a review.** *Afr Health Sci.* 2017; **17**(1): 225–236.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. World Health Organization: **Response to Measles Outbreaks in Measles Mortality Reduction Settings: Immunization, Vaccines and Biologicals.** 2009.
    PubMed Abstract

14. Rakesh P, Sherin D, Sankar H, *et al.*: **Investigating a community-wide outbreak of hepatitis a in India.** *J Glob Infect Dis.* 2014; **6**(2): 59–64.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. R Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria. 2017.
    Reference Source

16. Jombart T, Frost S, Nouvellet P, *et al.*: **outbreaks: A Collection of Disease Outbreak Data.** R package version 1.3.0. 2017.
    Reference Source

17. Almende BV, Thieurmel B, Robert T: **visNetwork: Network Visualization using 'vis.js' Library.** R package version 2.0.3. 2018.
    Reference Source

18. Lewis BW: **threejs: Interactive 3D Scatter Plots, Networks and Globes.** R package version 0.3.1. 2017.
    Reference Source

19. Fine PE: **The interval between successive cases of an infectious disease.** *Am J Epidemiol.* 2003; **158**(11): 1039–1047.
    PubMed Abstract | **Publisher Full Text**

20. Nöremark M, Widgren S: **EpiContactTrace: an R-package for contact tracing during livestock disease outbreaks and for risk-based surveillance.** *BMC Vet Res.* 2014; **10**: 71.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Carroll LN, Au AP, Detwiler LT, *et al.*: **Visualization and analytics**

tools for infectious disease epidemiology: a systematic review. *J Biomed Inform.* 2014; **51**: 287–298.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Guthrie JL, Alexander DC, Marchand-Austin A, *et al.*: **Technology and tuberculosis control: the OUT-TB Web experience.** *J Am Med Inform Assoc.* 2017; **24**(e1): e136–e142.
PubMed Abstract | **Publisher Full Text**

23. Nagraj VP, Jombart T, Randhawa N, *et al.*: **epicontacts (Version 1.1.1).** *Zenodo.* 2018.
http://www.doi.org/10.5281/zenodo.1210993

Open Peer Review

Current Peer Review Status: ❓ ✔ ✔

Version 2

Reviewer Report 14 June 2019

https://doi.org/10.5256/f1000research.18105.r49529

✔ **Josie Athens** (iD)

Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand

The authors present an R package that helps in the visualisation and analysis of epidemiological contacts.

It is a well-presented summary of the capabilities of the software and brings network theory tools to academics from the areas of public health and epidemiology.

For version 2 of the manuscript, the authors follow the Introduction with Use cases, which benefits the understanding of how data has to be prepared for input. I followed the examples, and the output that they present about the ebola data set is outdated. I suggest to include the version of the outbreaks package used in the examples. For those not familiar with R, presenting the data with the str command is not clear; to use the command head instead, could be a better option. This is something minor and not needed but would improve the understanding of the software on the most critical point, which is data input.

As in version 2, the Use cases section is presented before the Methods section, the last paragraph from the Use cases section needs to be changed from: "methods described above" to "methods described below".

The visualisations produced from epicontacts are impressive. For the manuscript to become more accessible to main stream readers, it would need to include the output from the analysis and interpretation.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Theoretical biology and biostatistics.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 02 August 2018

https://doi.org/10.5256/f1000research.15777.r36044

Peter Adebayo Adewuyi
Liberia Field Epidemiology Training Program, Monrovia, Liberia

This is a good software developed which could help in continuous visualization of contacts and their progression in disease tracking.

It is user friendly for those who are not computer specialist and still want to visualize data.

Data visualization is pertinent to disease monitoring and what the authors have done will aid in helping epidemiologist and public health specialist involved in outbreak response to quickly visualize progression and spread of disease from primary to secondary contacts and how the disease is evolving among contacts.

The software will actually achieve its purpose as stated in the conclusion of the write-up. Good work done.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 31 May 2018

https://doi.org/10.5256/f1000research.15777.r34084

?

Melissa A. Rolfes
Epidemic Intelligence Service (EIS), Influenza Division, Centers for Disease Control and
Prevention (CDC) , Atlanta, GA, USA

The article describes an R-based software tool aimed to facilitate analysis of data from outbreaks that include line lists of cases and case-contact data. The R package, epicontacts, is part of a larger suite of tools housed at the R Epidemics Consortium (RECON). The epicontacts package has the ability to merge data about cases in a line list with case-contact details, which then allows the user to describe and visualize contact networks, incubation periods, and serial intervals within an outbreak.

The codes and methods for analysis are partly described in the article, and the authors should provide a link to the packages documentation, either at CRAN or RECON webpages, where readers could learn more about the package and its options.

The output of the package provided in the article was interesting and intriguing. I felt that it was only partly explained and the article could benefit from the authors annotating the output and its interpretation a bit further. I have explored the RECON website and found the RECON Learn modules to be quite helpful in providing annotation of the epicontacts output and some guidance on interpretation. I would recommend that the authors consider either expanding the annotation of the output in this article or explicitly direct readers to the RECON Learn website for further instruction.

Additional suggestions:

- Consider moving the section of the article called "Use cases" to before the "Data handling" subsection of the "Implementation" section. I felt that the description of the input datasets under "Use cases" was very informative and would have been organizational more helpful had it been placed earlier in the article.

- Consider describing the sample outbreak data in a bit further detail. It appears to be data describing the MERS outbreak that occurred in South Korea in 2015. I think the description should include whether the data are simulated or from a real outbreak (if from a real outbreak, then a reference to the outbreak description should be included), the scenario of the outbreak, how many cases, how many contacts, place of the outbreak, duration of the outbreak, and a brief description of the demographic details included in the dataset. This amount of detail would allow the reader to translate the details of the outbreak from your text to the output provided by epicontacts.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research