

RECON

Building the next generation of statistical tools for outbreak response using R

Thibaut Jombart

18th April 2017

Imperial College London
MRC Centre for Outbreak Analysis and Modelling

Outline

Ebola response

Lessons learnt from the Ebola response



Lessons learnt from the Ebola response



Lessons learnt from the Ebola response

WHO Ebola response team

Help improving situation awareness

EBOLA OUTBREAK

December 2013 March 2014 August 2014 September 2015

First case WHO notified First data/report Latest data update

Imperial College Ebola team

Lessons learnt from the Ebola response



Most statistical/modelling tools for situation awareness missing.

What tools do we need?

Some examples:

- **data cleaning:** dictionaries, entry matching
- **graphics:** case incidence in space and time, contact tracing
- **parameter estimation:** key delays, transmissibility
- **estimate / test CFR:** gender, health care workers, treatments effects
- **predictions:** case incidence, mortality, evaluate interventions
- **report:** (semi-)automated situation reports

Who do we need to develop these tools?



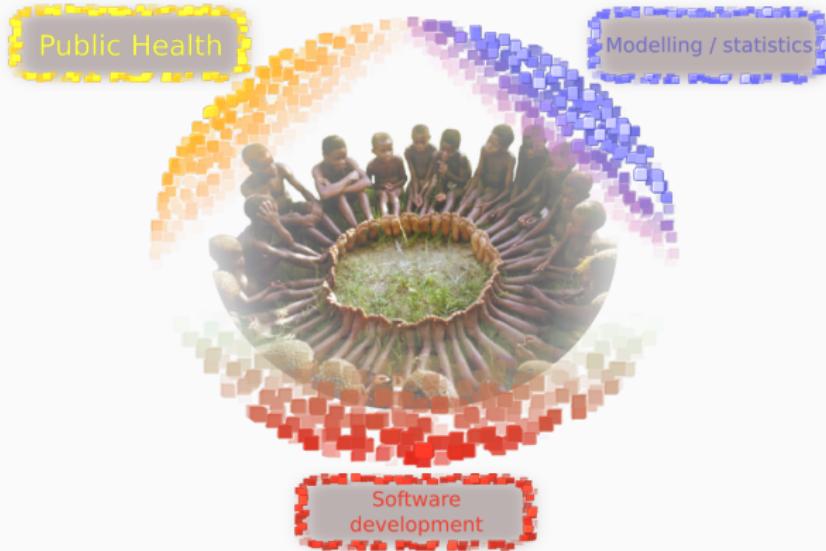
Who do we need to develop these tools?



Who do we need to develop these tools?



Who do we need to develop these tools?



The R Epidemics Consortium

Hackout 3: a hackathon for emergency outbreak response

Last summer at the *rOpenSci* headquarters (Berkeley)



Hackout 3: from ideas to projects to...



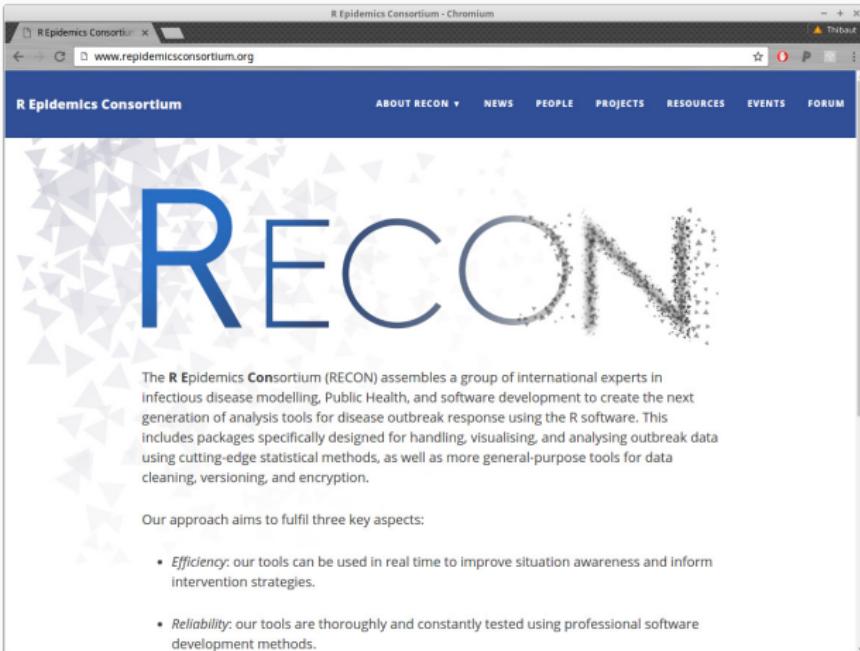
Hackout 3: from ideas to projects to...



How do we keep momentum once the event is over?

RECON: the R Epidemics Consortium

A taskforce to build a new generation of outbreak response tools in .



The screenshot shows a web browser window displaying the 'R Epidemics Consortium' website. The title bar reads 'R Epidemics Consortium - Chromium'. The address bar shows the URL 'www.repidemicsconsortium.org'. The page has a dark blue header with the text 'R Epidemics Consortium' on the left and navigation links 'ABOUT RECON', 'NEWS', 'PEOPLE', 'PROJECTS', 'RESOURCES', 'EVENTS', and 'FORUM' on the right. Below the header is a large, stylized word 'RECON' composed of numerous small, dark grey triangles. Underneath the word, there is descriptive text about the consortium's purpose and approach, followed by a section on key aspects and a bulleted list of requirements.

The R Epidemics Consortium (RECON) assembles a group of international experts in infectious disease modelling, Public Health, and software development to create the next generation of analysis tools for disease outbreak response using the R software. This includes packages specifically designed for handling, visualising, and analysing outbreak data using cutting-edge statistical methods, as well as more general-purpose tools for data cleaning, versioning, and encryption.

Our approach aims to fulfil three key aspects:

- *Efficiency*: our tools can be used in real time to improve situation awareness and inform intervention strategies.
- *Reliability*: our tools are thoroughly and constantly tested using professional software development methods.

www.repidemicsconsortium.org

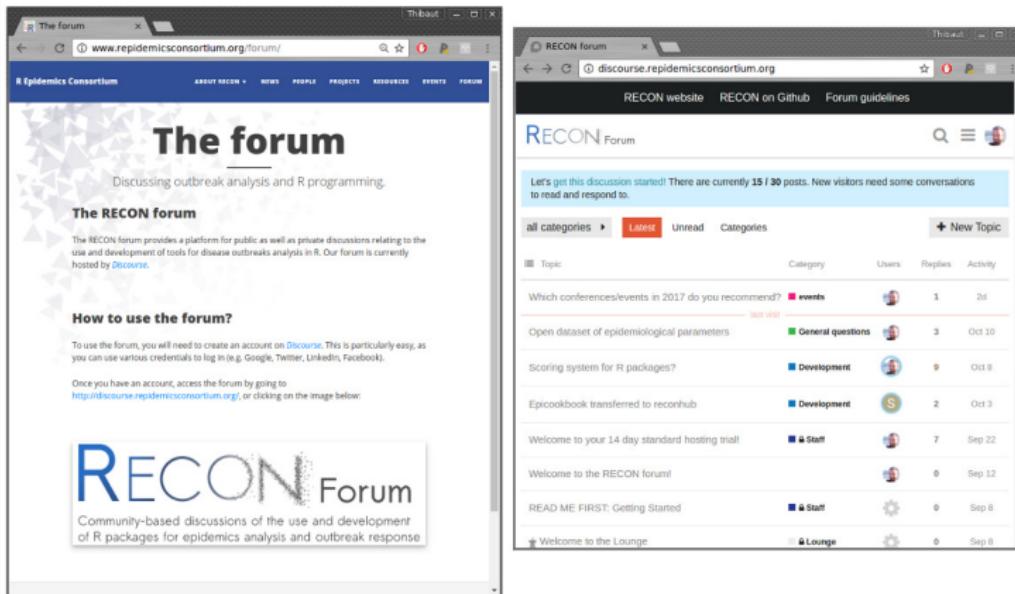
RECON

www.repidemicsconsortium.org

- started 6th September 2016
- 60 people (54 members, 6 board)
- 14 countries, > 30 institutions
- 2 new packages released, ~ 10-15 in development
- involvement in training programmes starting in 2017 (FETP, EPIET, ...)
- **public forum**, blog, online resources

The RECON forum

A platform for discussing epidemics analysis in .

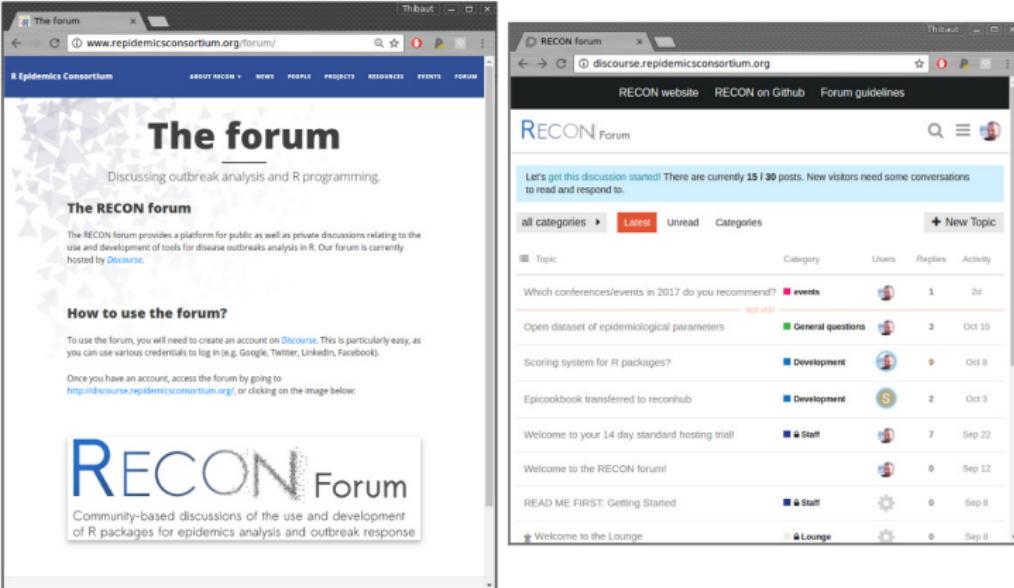


The image shows two side-by-side browser windows. The left window displays the official RECON forum at www.repidemicsconsortium.org/forum/. It features a dark blue header with the 'Epidemics Consortium' logo and navigation links for About RECON, News, People, Projects, Resources, Events, and Forum. The main content area has a light gray background with a geometric pattern and contains the heading 'The forum' and a sub-section titled 'The RECON forum'. The right window shows a discourse instance at discourse.repidemicsconsortium.org. The header includes links to the RECON website, GitHub, and Forum guidelines. The main area is titled 'RECON Forum' and shows a list of forum topics. A message at the top encourages users to start conversations. Below it, a table lists topics such as 'Which conferences/events in 2017 do you recommend?', 'Open dataset of epidemiological parameters', and 'Scoring system for R packages?'. Each topic includes a small profile picture, the number of replies, and the date it was posted.

www.repidemicsconsortium.org/forum

The RECON forum

A platform for discussing epidemics analysis in .



The RECON forum provides a platform for public as well as private discussions relating to the use and development of tools for disease outbreaks analysis in R. Our forum is currently hosted by [Discourse](#).

To use the forum, you will need to create an account on [Discourse](#). This is particularly easy, as you can use various credentials to log in (e.g. Google, Twitter, LinkedIn, Facebook).

Once you have an account, access the forum by going to <http://discourse.repidemicsconsortium.org/>, or clicking on the image below:


Community-based discussions of the use and development of R packages for epidemics analysis and outbreak response

www.repidemicsconsortium.org/forum

Join us!

RECON package: what do we aim for?

- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**

RECON package: what do we aim for?

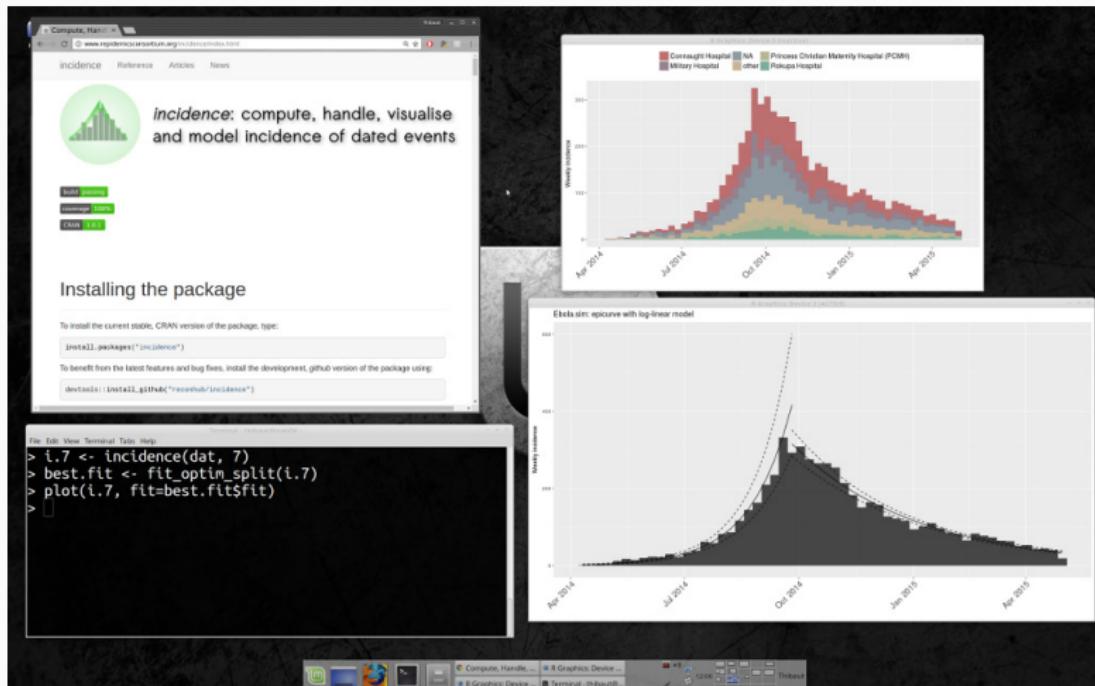
- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**
- **reliability**: outputs can be trusted; **continuous integration, extensive unit testing, code review, good practices**

RECON package: what do we aim for?

- **efficiency**: useful for improving situation awareness in real time; **cutting-edge, computer-efficient statistical methods**
- **reliability**: outputs can be trusted; **continuous integration, extensive unit testing, code review, good practices**
- **accessibility**: widely available, easy learning curve; **extensive documentation, tutorials, websites, forum**

Up-and-coming RECON packages

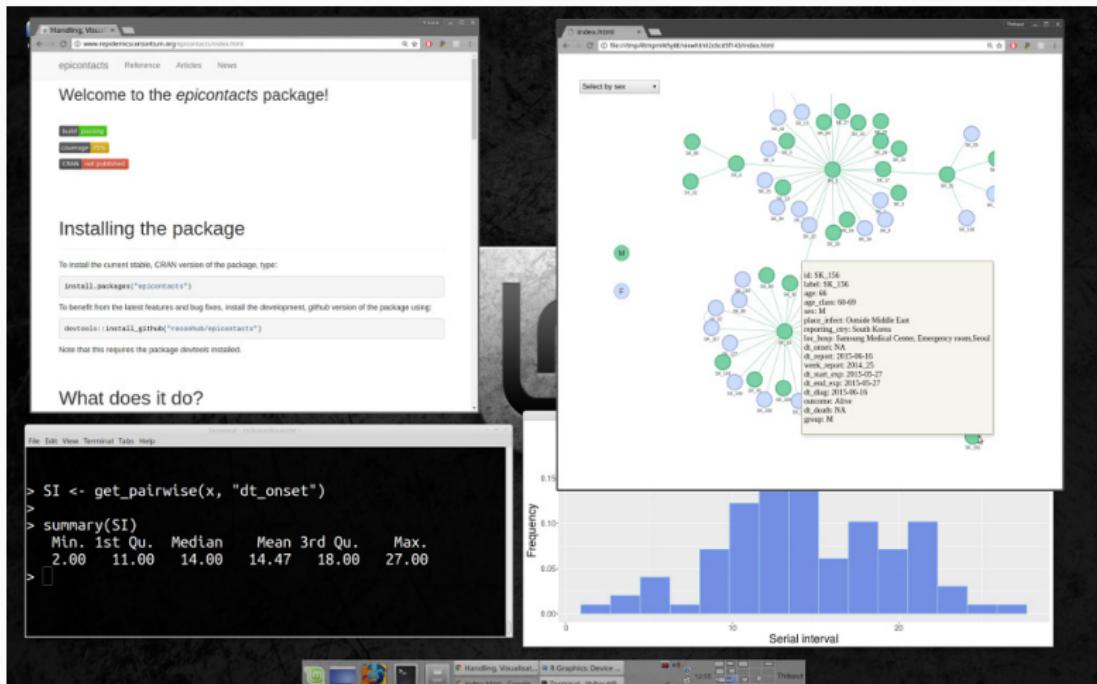
incidence: computation, handling, visualisation and modelling of epicurves



www.repidemicsconsortium.org/incidence

[released]

epicontacts: handling, visualisation and analysis of epidemiological contacts



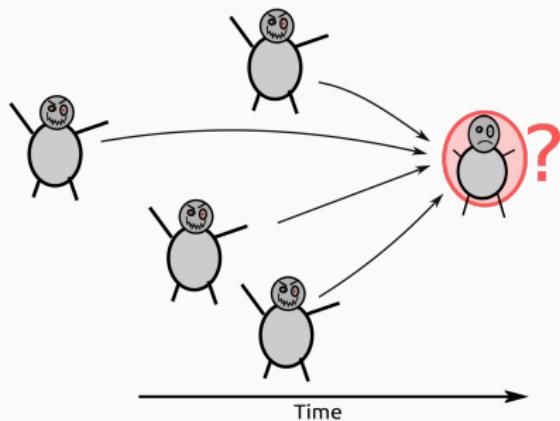
www.repidemicsconsortium.org/epicontacts

[release May 2017]

outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

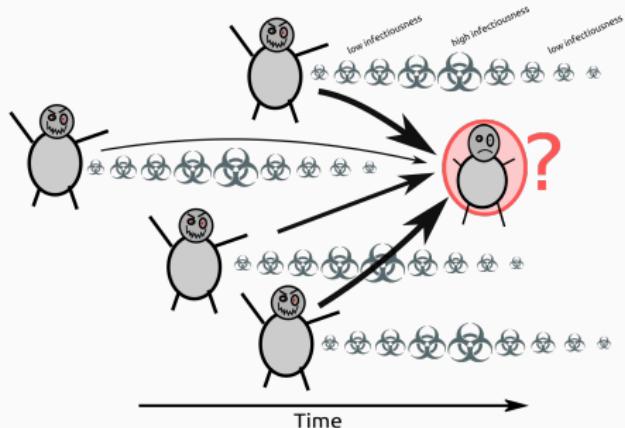
(Jombart et al, PLoS Comp Biol, 2014)



outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

(Jombart et al, PLoS Comp Biol, 2014)

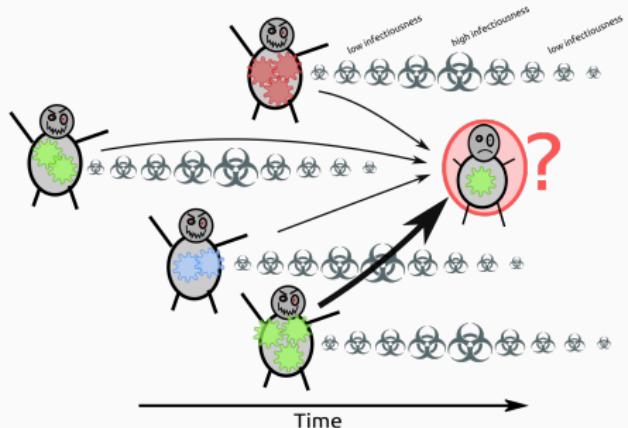


Since *outbreaker*: new models, data, and questions.

outbreaker2: inferring who infects whom in an outbreak

Original *outbreaker* model: timing of symptoms and pathogen genomes to infer infectors

(Jombart et al, PLoS Comp Biol, 2014)



Since *outbreaker*: new models, data, and questions.

But: methodological niche fragmented.

Are different methods really... different?

Are different methods really... different?



Are different methods really... different?



Different models can lead to very similar implementations.
Can we find a **general formulation**?

What do these model look like?

- a, b, c : different types of data
- θ : parameters / augmented data

Data are often assumed to be *conditionally independent*:

$$p(a, b, c|\theta) = p(a|\theta)p(b|\theta)p(c|\theta)$$

What do these model look like?

- a, b, c : different types of data
- θ : parameters / augmented data

Data are often assumed to be *conditionally independent*:

$$p(a, b, c|\theta) = p(a|\theta)p(b|\theta)p(c|\theta)$$

Components can be treated as **plugins**.

outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.

outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.

outbreaker2: a general cauldron for cooking methods

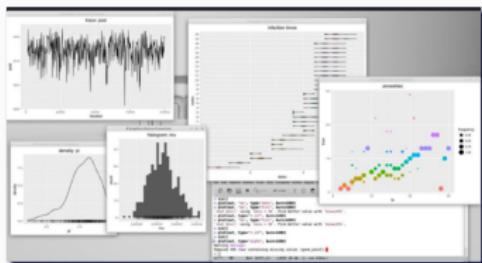
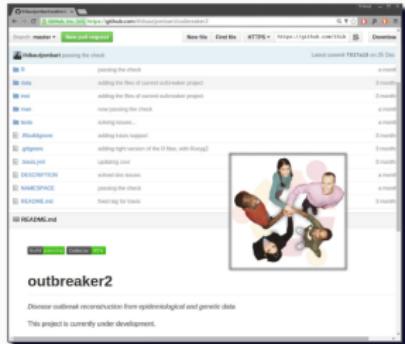
Use-your-own: data type, likelihood, prior, MCMC.

outbreaker2: a general cauldron for cooking methods

Use-your-own: data type, likelihood, prior, MCMC.

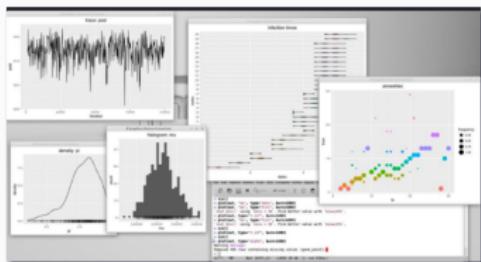
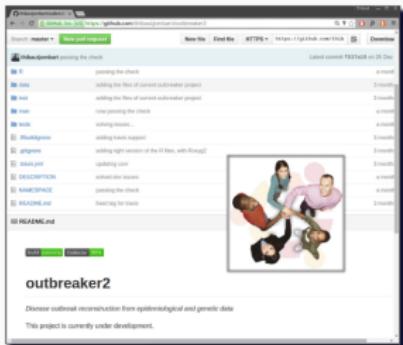
Modularity is key to generalising approaches

outbreaker2: a general tool for outbreak reconstruction

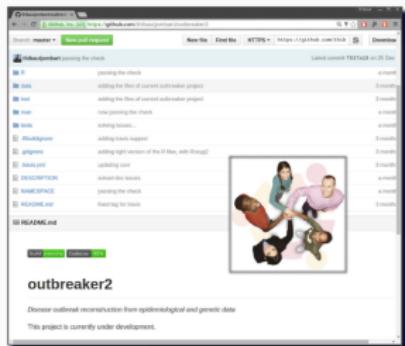


outbreaker2: a general tool for outbreak reconstruction

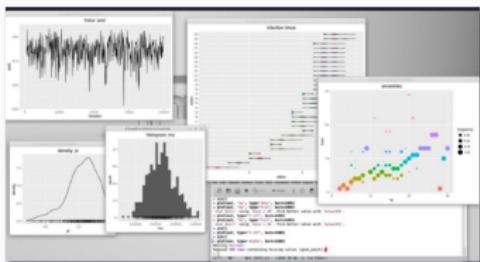
- **modularity:** likelihood, priors, samplers are all modules



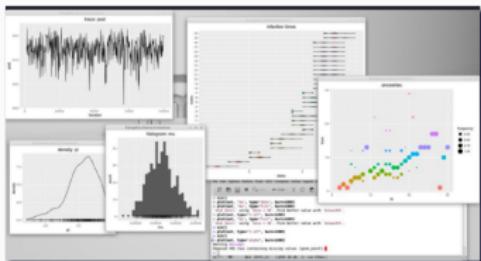
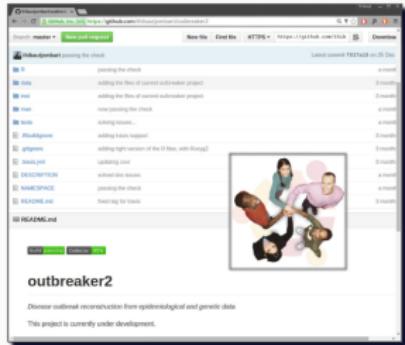
outbreaker2: a general tool for outbreak reconstruction



- **modularity:** likelihood, priors, samplers are all modules
- **new 'extensions':** contact tracing, spatial structure, new MCMC

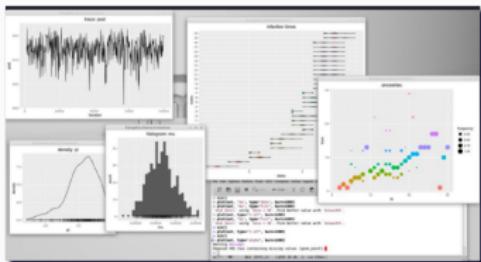
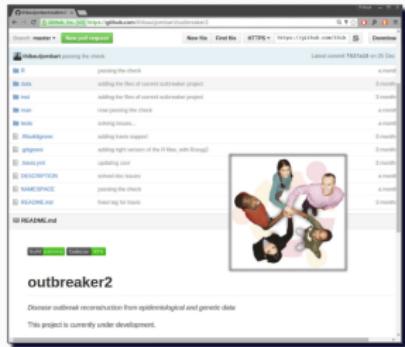


outbreaker2: a general tool for outbreak reconstruction



- **modularity:** likelihood, priors, samplers are all modules
- **new 'extensions':** contact tracing, spatial structure, new MCMC
- **reliability:** continuous integration, extensive unit testing (aiming for 100% coverage)

outbreaker2: a general tool for outbreak reconstruction



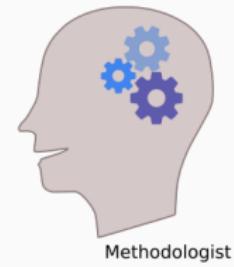
- **modularity:** likelihood, priors, samplers are all modules
- **new 'extensions':** contact tracing, spatial structure, new MCMC
- **reliability:** continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier:** plot methods using *ggplot2*, interactive networks visualisation

outbreaker2: a general tool for outbreak reconstruction

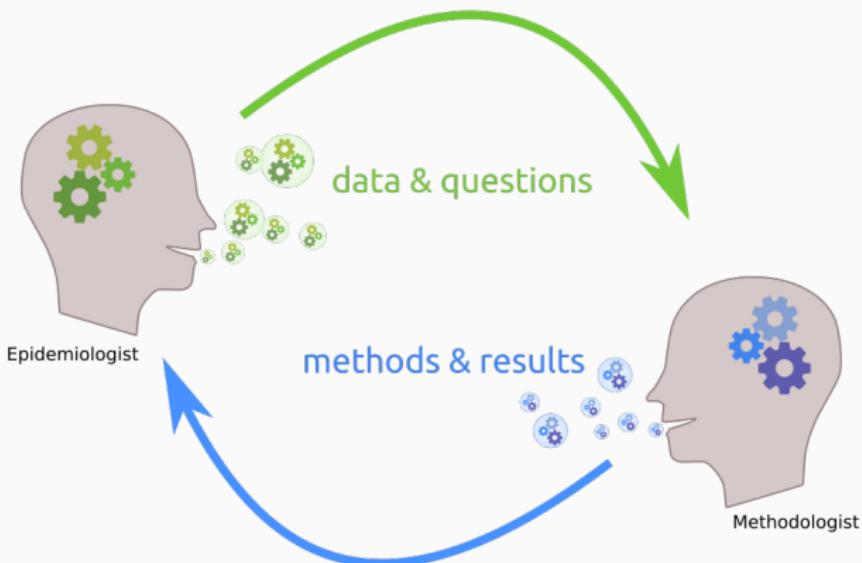
- **modularity**: likelihood, priors, samplers are all modules
- **new 'extensions'**: contact tracing, spatial structure, new MCMC
- **reliability**: continuous integration, extensive unit testing (aiming for 100% coverage)
- **prettier**: plot methods using *ggplot2*, interactive networks visualisation
- should **facilitate new contributions**

Methodological dialogue

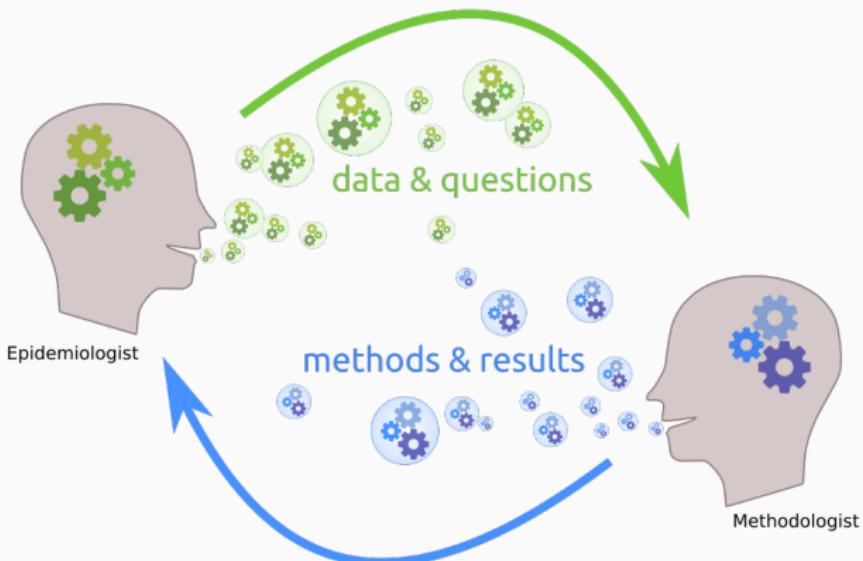
Methodological development relies on an interdisciplinary dialogue



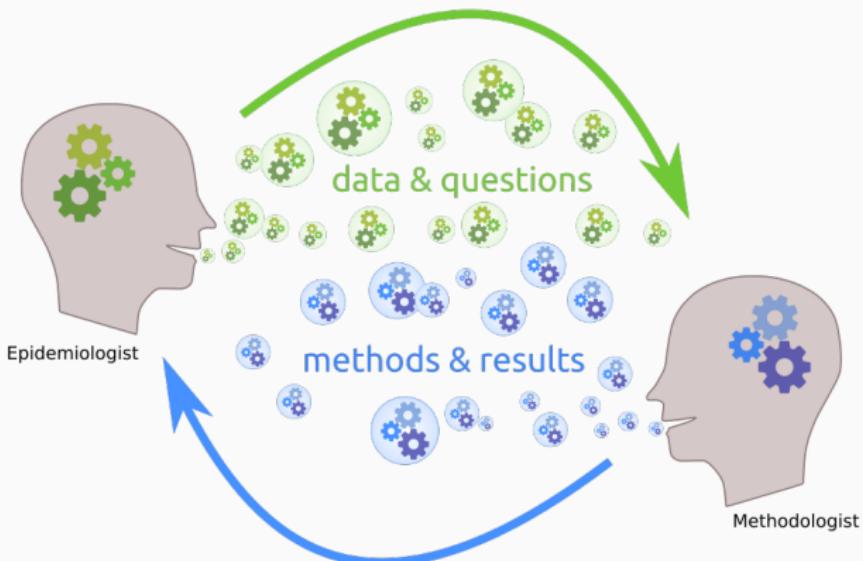
Methodological development relies on an interdisciplinary dialogue



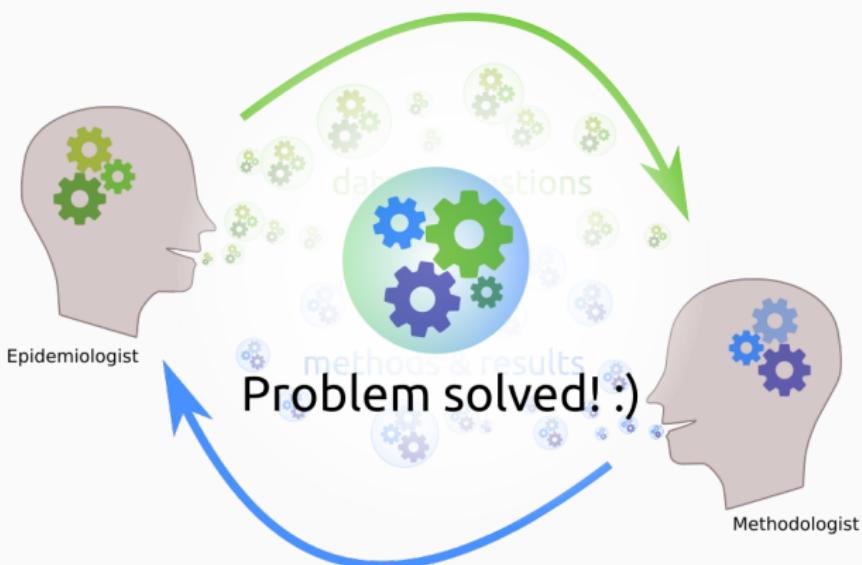
Methodological development relies on an interdisciplinary dialogue



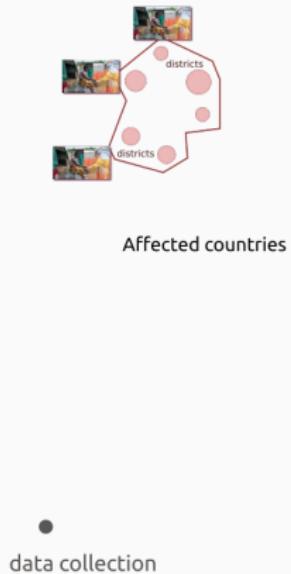
Methodological development relies on an interdisciplinary dialogue



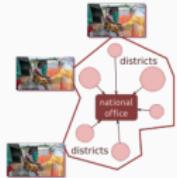
Methodological development relies on an interdisciplinary dialogue



Outbreak response context creates distance and delays



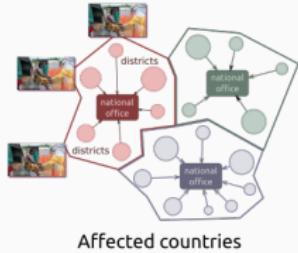
Outbreak response context creates distance and delays



Affected countries

time (block = day)
● 
data collection

Outbreak response context creates distance and delays



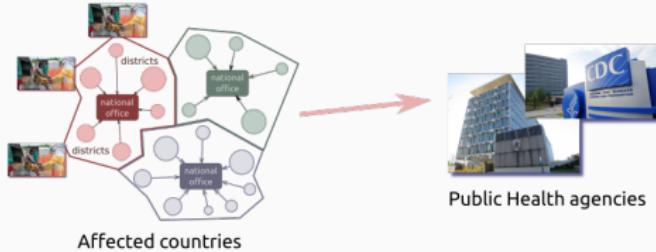
Affected countries

time (block = day)

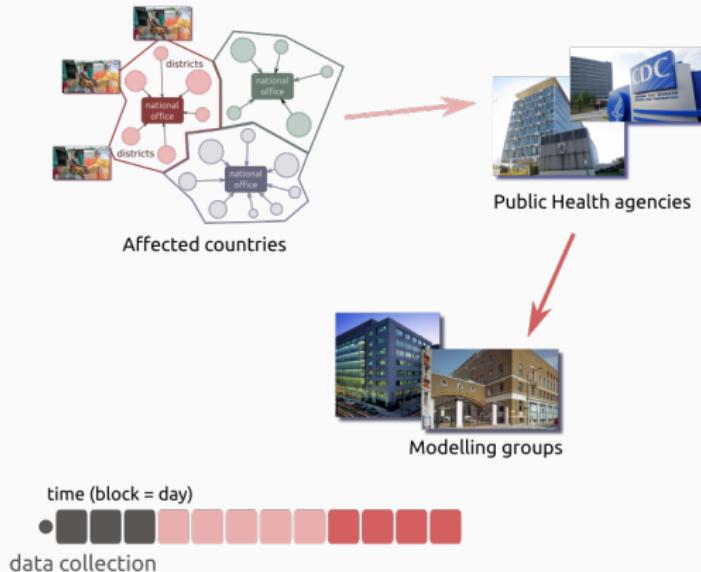


data collection

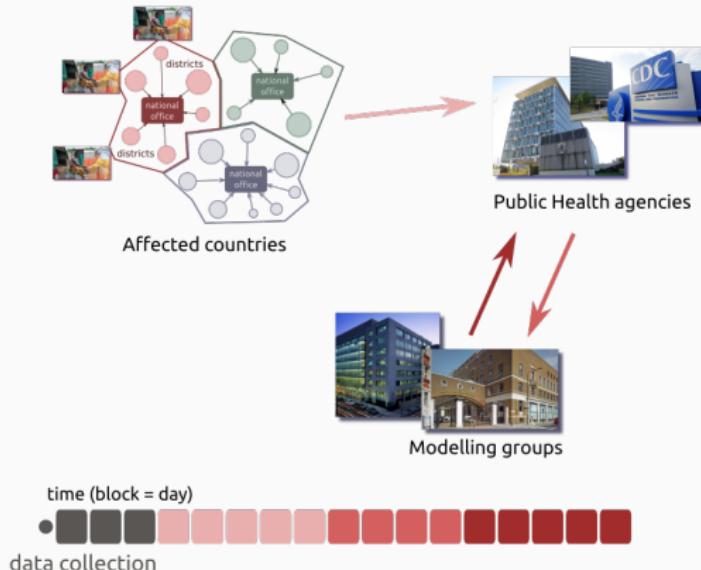
Outbreak response context creates distance and delays



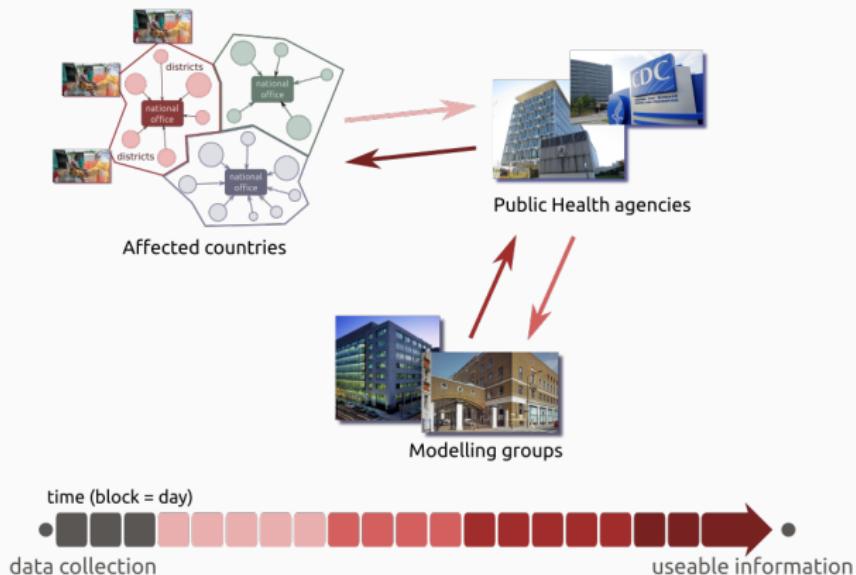
Outbreak response context creates distance and delays



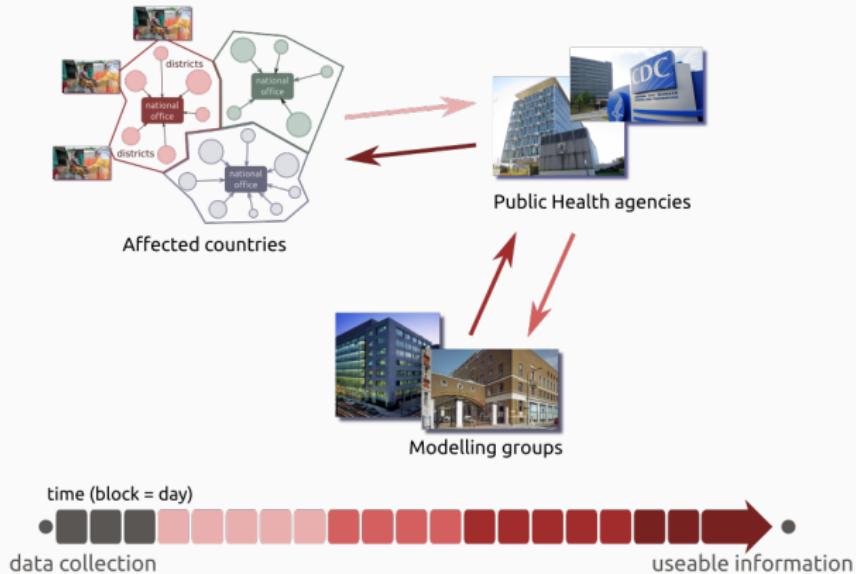
Outbreak response context creates distance and delays



Outbreak response context creates distance and delays

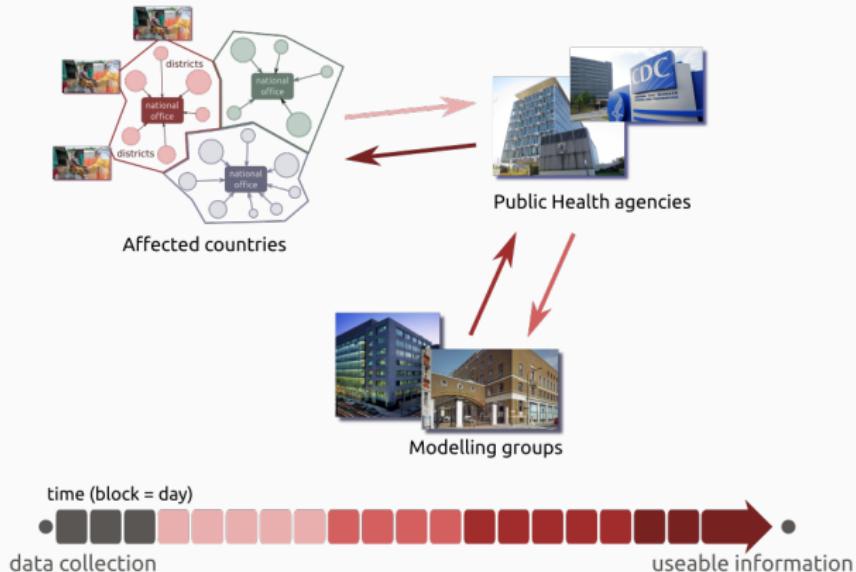


Outbreak response context creates distance and delays



- efficient tools can shorten delays

Outbreak response context creates distance and delays



- efficient tools can shorten delays
- potential of **embedding methodologists in outbreak response teams**

Thanks to...

- **Mirna Panic**
- **Imperial College:** Neil Ferguson, Rich Fitzjohn, Anne Cori, Finlay Campbell, Evgenia Markvardt, James Hayward
- **UC Berkeley:** Karthik Ram
- **Groups:** WHO Ebola Response Team, Hackout 1/2/3, RECON members, GOARN
- **funding:** HPRU-NIHR, MRC

More on:

www.repidemicsconsortium.org

Questions?