Plague Outbreak in India

Challenge Assignment

December 16, 2019

Situation

There is a pneumonic plague outbreak occuring in northeast India. The community where the outbreak is occurring has a population of 152, and so far there have been a total of 62 suspected, probable, and confirmed cases.

Data set

You are provided with a line list that includes the following data:

- Case ID - unique identifier for each individual case
- Household ID - unique identifier for each household in the community
- Date of onset - the date of onset of symptoms
- Date of death - the date the patient died; NA if the patient recovered and/or if the patient is still ill
- Date of report - the date that the case was notified
- Status of case - suspected, probable, or confirmed

Assignment

You are asked to write a report on the outbreak to date based on the information in the line list. Working with your team, you should decide what questions the report will answer, but as a *minimum* it needs to include an informative epidemic curve of the outbreak so far. You may use any tools available to you to write the report, but the generation of the report should be automated. You should provide a link to the code repository and provide read access to `jrcpulliam` at the time of submission. (NB: Assuming the code repository is on GitHub or GitLab. If it's elsewhere, please send the zipped repo as an attachment with your report submission.) Make sure to include a README file in the repository with information on the workflow that generates the report. The report itself should also be emailed to `pulliam@sun.ac.za` as a `.pdf` or `.html` file by the deadline given below.

In addition to your report, please include in your email a description of *two* types of additional data that you would like to have that are not included in the line list. Justify your choices.

Deadlines

- **Initial report:** 2020-01-17
- **Final report:** 2020-02-13

  NB: Following submission of the initial report, you will be given feedback on the report and updated data on the outbreak. You may also receive additional types of data if your justification is good and the data you've requested are available. Note that neither genome sequence data nor GPS locations will be available, though you still might want to make an argument for them.

Reminder

You are asked to spend 4 hours per week on SMART activities, *with the exception of 15 December - 1 January*. You are likely to need the full 8-10 hours before the initial deadline to generate your report; if you don't, you are encouraged to continue working with the self-learning resources or other related tasks. Also, if you *do* want to work on SMART activities during the break period, please be mindful of your group members' time and ensure that everyone is able to make a substantial contribution to the report. This may mean working on *other* SMART-related activities through the end of 2019 to accomodate their schedules. It would be good to communicate with them ahead of time about availability.