## Simulated Outbreak Exercise

|  |  |
| --- | --- |
| Scenario Date: | Wednesday 11 October (09:15-12:30, including coffee break) |
| Inject No: | 11 |
| Inject length: | 2 hours 45 minutes |
| Inject Title: | Describe the cases by person, place (not here!) and time |

1. **Learning outcomes:**At the end of the session, participants will be able to:

- Discuss which descriptive statistics would you use to describe the cases,  
- Describe the cases by person and place (this is an outbreak in a high school, so we won’t explore “place” this time!).  
- Generate preliminary hypotheses bases on the descriptions.

Facilitator’s note: This is the longest R-inject. Consider realising solutions sooner than hal-way throught the inject, depending on the mood of your group. Let them know in advance that this is the longest R-inject.

1. **Story/plot description:**  
   You have finished cleaning your dataset and will now use it to describe the people identified as cases so far and generate a hypothesis to inform the next steps in the outbreak investigation.

Note that you should try to have a data analysis plan *before* any data collection. Nevertheless, during the urgency of outbreak investigations, questionnaire development may sometimes occur before you had time to draft a proper analysis plan (as it is the case in this simulated outbreak).

Drawing “dummy” tables and graphs will help you precise the type of statistics you will carry out later on.

Facilitator’s note: if needed, check with fellows about where are we in the outbreak 10-step framework, so everyone is aware of the process, and no one gets lost in the R code.

1. **Questions/assignments for the group:**

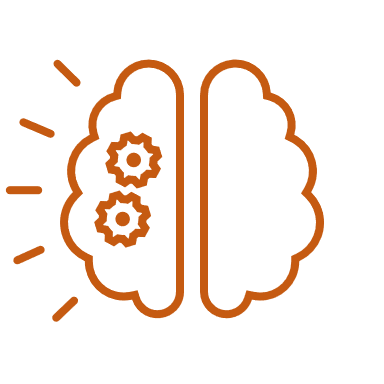
If you closed your R project, please open it again and load the clean dataset. We recommend you use the “Copenhagen\_clean2.rds” provided by us in EVA. This is just to be sure all needed changes in that file for the code to work have been done (as you may have changed the file in a slightly different way than we have).

3.1. Install packages (if needed) and load libraries.

3.2. Import your data.

3.3. Describe the outbreak in terms of **time** using ggplot2. To learn more about ggplot2 you can check [The Epidemiologist R Handbook: Chapter on **ggplot**](https://extranet.ecdc.europa.eu/Training/PM/Documents/Module%20Preparations/Intro%20Course/Track%20-%20Simulated%20Outbreak/Injects/The%20Epidemiologist%20R%20Handbook:%20Chapter%20on%20ggplot) and [ggplot2 - Elegant Graphics for Data Analysis](https://ggplot2-book.org/index.html).

1. Create a histogram (geom\_histogram is a good function to use for this!) to visualize the incubation period.



**Stop and reflect**: What can you infer from the incubation periods in the histogram?

Facilitator’s note: Recommend fellows to reflect on lecture given on first week by Pawel on epicurves. This looks a quite short incubation time!

1. Create an epicurve for the date and time of onset (using onset\_datetime), limiting the input data to cases.
   * Use scale\_x\_datetime() and set date\_breaks .
   * Label your x and y axes using labs.



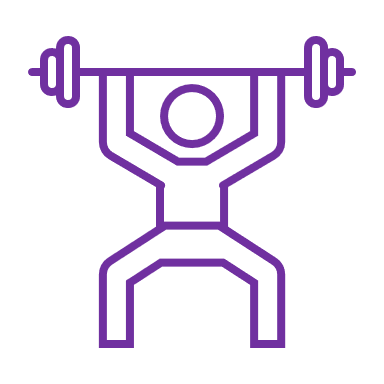
**Hint:** A rule of thumb is to use one third or one fourth of the average incubation period as an interval. For our investigation this means we should use approximately a 6h interval for the date\_breaks argument.

1. Building up on your previous epicurve, create another epicurve to compare between sexes and additionally investigate how teachers versus students were distributed.
   * Use fill = group
   * Use sex as your facets for facet\_wrap()

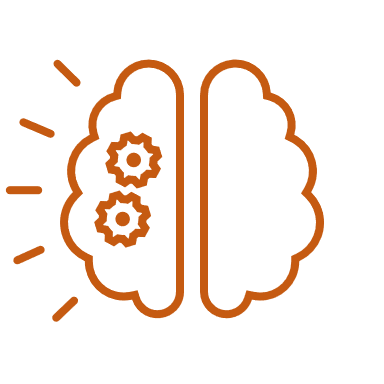


**Hint:** fill adds an additional variable to be displayed in the plot: group is going to determine the fill-colour of our bars.

facet\_wrap splits the graph in two: one each for the two levels of sex.



**Extra**: If you want, try using str\_glue() to add the total number of cases to the sub-title of the plot. str\_glue() is a very useful function that allows you to dynamically create a summary statistic from your data within some normal text.



**Stop and reflect**: What does the stratified epicurve tell you? Does the shape of the epicurve support a viral or toxic aetiology? What other information can you obtain from it?

Facilitator’s note: When constructing an epicurve, we need to decide on the resolution, i.e. the time interval for a single bar. A rule of thumb is to use one third or one fourth of the average incubation period as an interval. For our investigation this means we should use approximately a 6h interval.

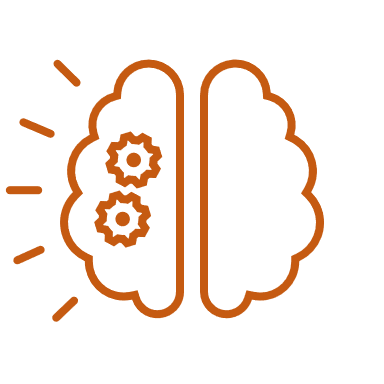
This seems a good choice, as we saw that the daily interval was too coarse to really see the signal we are after. The epicurve and the summary of the incubation period show that there seemed to be a rapid onset of symptoms following exposure. This is in line with our previous suspicion that a virus or a toxin might be the causative agent in the outbreak.

The unimodal shape with the sharp peak suggests a point source, while the tail on the right-hand side could be explained by secondary cases or background noise. Also, people that only consumed a little contaminated food and therefore only a low infectious dose could have a longer incubation period and could explain the late cases.

The above results are in line with norovirus as the prime suspect, but the symptoms are not a textbook fit. There are too few people that experienced vomiting! Looking forward to receiving the lab results!

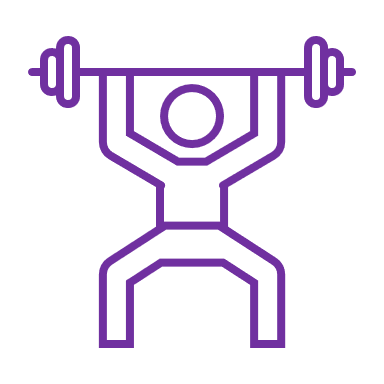
3.4. Describe the outbreak in terms of **person** ; you could use tabyl(). For more detail on tabyl() function and the adorn\_XYZ helpers, see the [janitor documentation](https://cran.r-project.org/web/packages/janitor/vignettes/tabyls.html) or [The Epidemiologist R Handbook section on tabulation](https://epirhandbook.com/en/descriptive-tables.html#tbl_janitor).

1. Cross-tabulation of cases with group
2. Create a Cross-tabulation of cases with sex.



**Stop and reflect**: What do you think of these results?

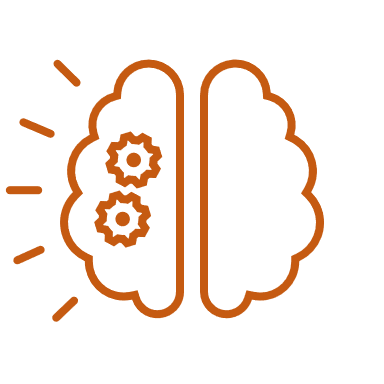
Facilitator’s note: The distribution of the cohort regarding sex, group and class also didn’t reveal anything unusual. Students seem a bit more affected by the outbreak than teachers and the attack rate is higher for older students in higher classes. This, however, is a purely descriptive result.



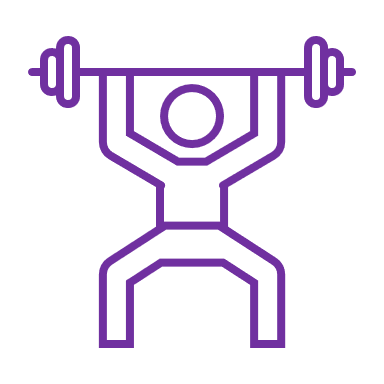
**Extra**: If you want, create an age-sex pyramid using the apyramid package. To do so, first create an age category variable with the epkit function age\_categories().

3.5 Explore the distribution of all clinical signs (symptoms).

1. Create a summary table with symptoms stratified by case definition, and present an overall column as well. You could use tabyl() or gtsummary::tbl\_summary() ; you can find further information about gtsummary in the [Epidemiologist R handbook section on gtsummary](https://www.epirhandbook.com/en/descriptive-tables.html#tbl_gt).



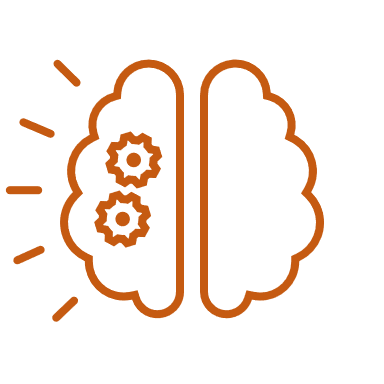
**Stop and reflect**: Do you think the symptoms selected for the case definition were the right ones, or would you change anything?



**Extra**: If you want, present the symptoms in an ordered bar chart. To do so, reshape the data using pivot\_longer(), and then group\_by(), summarise() and count() to tally up the counts for each symptom, stratified by case definition. Use ggplot2::coord\_flip() for a better visualisation.

3.6 Attack proportions

1. Calculate the overall attack proportions (percentage of cases among the total observed individuals). You could use tabyl().



**Stop and reflect**: What can you infer from the overall attack proportion about this outbreak and possible vehicles/exposures?

Facilitator’s note: The overall attack proportion is 57.3%. This means that more than half of the people who ate a meal were cases!

1. Calculate attack proportions for group, class and sex by case status. You could use tabyl() (as you did in 3.4) or gtsummary::tbl\_summary.

3.7 Draft the relevant paragraph(s) and table(s) and/or figure(s) in your outbreak report (Methods and results).



**Hint:** To save your tabyl table as a .docx, you could convert your tabyl to flextable (as\_flex\_table()), ensure only one line per row (flextable::autofit()) and save as .docx (flextable::save\_as\_docx(path = "nameoftable.docx")

1. **Notes for facilitators:**Some fellows may still struggle with the R interface. Remind them that they always need to repeat the same steps by creating/using the workspace with access to the correct folders (R project) and that they must load the packages they will use and import the right dataset.
2. **Scenario for the role play:**N/A
3. **Timing:**2 hours 45 minutes
4. **Need for materials (logistics)**N/A
5. **Deliverables**A script with reproducible code and a paragraph in the Methods section summarizing the approach to the descriptive analysis, AND the completed section in the Results including tables, graphs and introductory text.