In your country, Shiga-toxin producing *Escherichia* *coli* belonging to serotype O157:H7 are one of the major causes of haemolytic-uremic syndrome. Strains collected from human cases are routinely sequenced and submitted to NCBI.

You have been asked to analyse the genomic data obtained between 2016 and 2020 (n = 432) and describe their main genomic features and the evolution of the population of this serotype.

You are provided with a metadata table containing the name, the Biosample accession, the Illumina SRA accession, and the isolation date of each strain (STEC\_dataset\_GenEpi-BioTrain.xlsx).

Using any of the tools presented during the course, try to answer the following questions about the dataset:

1. Were there outbreaks during the evaluated period? How many? Present how they were identified.
2. What are the main virulence determinants present in the isolates? In particular, which is the main Shiga-toxin type found among them?
3. Did you observe any temporal trend during the study period (i.e., seasonality, variation in the number of strains over the years, etc.)?
4. Even though antimicrobial chemotherapy is not recommended for the treatment of STEC infections in many countries, can you identify AMR genes among the studied population, particularly acquired genes?