BIO319 assessment 2 (Data cleaning)

1. Setting up your project

As for the workshops, accept the assignment invite on GitHub classroom, then create a project in PositCloud within the BIO319 workspace. The project contains lots of .txt files, each a different dataset for each student. Identify your dataset by your student ID in the file name. You can then delete the other datasets.

2. The assignment

The dataset is untidy and your assignment is to inspect the data, identify how it is untidy, and then write code to tidy it.

We expect you to produce three files, as described below. **Make sure the files have suitable names and have your student ID in the file name.**

1) A text file that describes the data set, i.e. which variables it contains, and it what ways it violates the rules of tidy data. You also need to justify the steps you have taken to tidy the data. This should have a maximum of 600 words, but can be shorter.

2) An R script that contains your code.

3) The cleaned data exported in a suitable format.

We will assess you work using the following criteria:

1) Text file: Does it describe the data correctly and identify the untidy aspects of the data? Is the tidying approach well justified? (25/100 marks)

2) Code file: Can we run the script as it is without error messages? Is it well annotated and properly formatted? Does it use succinct, but readable code, avoiding overly complicated solutions? (45/100 marks)

3) Data file: Is the output clean and identical to the output when running your code? Is the file in a suitable format? (30/100 marks)

3. Assignment submission

The assignment is released Monday 6th March 18:00 via email and due Friday 24th March 16:00. You will submit your assignment by pushing the files to your GitHub repository.