Practice question 3

The dataset '*lichen.csv*' is available on Moodle. Different species of lichen are often identified as tolerant or intolerant to atmospheric pollution, and hence the occurrence of different lichen species has been used as a guide to pollution levels. This dataset contains the results of a study looking at the lichen growth rate of a single species at various locations across the UK. At each location several measures of pollution were recorded as well as the number of bacteria on each sampled lichen. The researchers are interested to know if the growth rate of an individual species can be related to the amount of pollution and if the growth rate is related to the bacterial community inhabiting the lichen. The variables in the dataset are;

plot – individual number of each sampled location.
growth_rate – lichen growth rate (cm^2/yr).
pollution_index – composite measure of pollution (no units), higher values indicate greater level of pollution.
NOx – pollution measurement, oxides of nitrogen (g/ha/yr).
bacteria – number of bacteria (measured in colony forming units) from sampled lichen branch (cfu/g).
NH3 – pollution measurement, atmospheric ammonia (ppb).

Examine the dataset to answer the following research questions;

1) Is there correlation between any of the measured variables? If so, statistically report the three strongest correlations.

2) Is there a relationship between lichen growth rate and the pollution index? [hint: think carefully about this relationship, does one variable directly impact the other?]

Marking guidance
2 marks for testing correlations; 3 marks for correctly reporting the correlations; 2 marks for testing the relationship between growth rate and pollution index; 2 marks for reporting the results statement; 1 mark for plotting the relationship, including a fitted/modeled line.

P.S. there will be no hints in the unit assessment.