The data we have chosen to use is a Stats NZ dataset on modelled lake water quality. There are 3820 lakes in New Zealand that have a surface area larger than one hectare. Of these, only 18 lakes had incomplete observations, so 3802 lakes are included in the sample. No lakes in offshore islands of New Zealand are included in this dataset. For each lake, modelled median values for the period 2013 to 2017 were recorded for each of the six measures; Ammoniacal Nitrogen, Chlorophyll-a, Total Nitrogen, Total Phosphorus, water Clarity and Trophic level index. Measures of the perimeter (in metres), depth (in metres) and the area (in square metres) were recorded as well as the ID of the lake, the region it is in, the dominant land cover and the lake type.

Ammoniacal Nitrogen is a form of nitrogen that supports algae and plant growth, but in large concentrations can be toxic to aquatic life. This is measured in milligrams per litre. The national bottom line for this measure is 1.3mg/L. No measure of Ammoniacal Nitrogen exceeded this so the variable indicating whether the lake had exceeded the national bottom line was deleted. Chlorophyll-a is an organic molecule found in plant cells that allows plants to photosynthesize. The variable Chlorophyll-a is a measure of the concentration of phytoplankton biomass in milligrams per cubic metre. High concentrations of chlorophyll is a symptom of degraded water quality. The national bottom line for this measure is 12 and many of the observations exceeded this. Clarity is measured in Secchi depth. This is the maximum depth (in metres) a black and white Secchi disk is visible from the surface of the lake. Total Nitrogen is the sum of all nitrogens found in the water, including organic nitrogen from plant tissue. An excess of nitrogen in lakes can cause an increase in algae and plant growth, possibly depriving the lake of oxygen. Total Nitrogen is measured in milligrams per cubic metre and the national bottom line for stratified lakes is 750mg/m^3, and for polymictic lakes is 800mg/m^3. There are observations exceeding the national bottom line. Total Phosphorus is the sum of all phosphorus forms in the water, including phosphorus bound to sediment. Large amounts of phosphorus in lakes can reduce dissolved oxygen in the water. This can cause low oxygen areas in the lake, where some aquatic life cannot survive. Total Phosphorus is measured in milligrams per cubic metre and has a national bottom line of 50mg/m^3. Trophic Level Index 3 rating (TLI3) is a composite index developed by Burns et al (2000) which describes the state of nutrient enrichment, comprising total nitrogen, total phosphorus and chlorophyll-a. Along with the rating, there are 5 categories, known as trophic states, based on the rating. These are; microtrophic (TLI3 < 2) indicating the lake is very clean, oligotrophic (TLI3 between 2 and 3) indicating the lake is clear with low concentrations of nutrients and algae, mesotrophic (TLI3 between 3 and 4) indicating the lake has moderate concentrations of nutrients and algae, eutrophic (TLI3 between 4 and 5) where the lake is murky and has high concentrations of nutrients and algae, and hypertrophic (TLI3 > 5) when the lake has extremely high concentrations of nitrogen and phosphorus and is very fertile.

The regions in this dataset are; Auckland, Bay of Plenty, Canterbury, Gisborne, Hawke's Bay, Whanganui, Marlborough, Northland, Otago, Southland, Taranake, Tasman, Waikato, Wellington and West Coast. Each lake corresponds to the region it is located in. There are four types of dominant land cover; Exotic Forest, Native, Pastoral and Urban area. There are 12 lakes with no entry for dominant land cover, however in the description of the dataset by Stats NZ, it states all lakes have been categorised, and indicated these empty entries should be another category called 'other' that includes 'Gorse and/or Broom', 'Surface mines and dumps', 'Mixed exotic shrubland', and 'Transport infrastructure' so we have assigned these to the other category. The category Urban area is applied if urban cover exceeds 15 percent of catchment area. Pastoral is applied if pastoral exceeds 25 percent of catchment area and not already assigned urban. The other three categories; Exotic forest, Native, or Other were assigned according to the largest land cover type by area, if not already assigned urban or pastoral. There are two types of lake based on the National Objectives Framework; Stratified and Polymictic. Polymictic lakes are too shallow to develop thermal stratification (the tendency to form separate thermal layers in warm weather). Stratified lakes develop thermal stratification. There are four band classifications, according to the National Objectives Framework. These are A, B, C and D, with A being the best state and D the worst. The boundary between bands C and D is the national bottom line for the relevant measures (ammoniacal nitrogen, chlorophyll-a, nitrogen and phosphorus).

We have edited this data slightly. Previously, Ammoniacal Nitrogen, Chlorophyll-a, Total nitrogen, Total phosphorus, Water clarity and Trophic level index were categories in a variable called 'measure', and the values for each were all recorded in a variable called 'median'. We wanted to be able to compare levels of some of these variables to see if there was a relationship between any pairs. To be able to do this, we made a new variable for each measure, with their values recorded in their own variables, along with some categorical variables indicating the band classification and whether the observation exceeded the national bottom line for that measure (if there is one). In this new set of data, each lake has a single observation, as opposed to the six or more each had before the data was truncated. No observations were changed and all 3802 lakes had observations for each of the six measures. We have also deleted variables indicating the New Zealand map grid coordinates of each lake, the period the data was modelled for (all were from 2013 to 2017) and the name of the lakes, as very few were named.

Our motivation is to explore the relationships between the measures of water quality and the location, land cover and type of New Zealand lakes.

References

Chlorophyll:
https://sciencing.com/what-are-the-roles-of-chlorophyll-a-b-12526386.html

https://www.epa.gov/national-aquatic-resource-surveys/indicators-chlorophyll

Data:
https://www.stats.govt.nz/indicators/lake-water-quality

https://data.mfe.govt.nz/table/104560-modelled-lake-water-quality-2013-2017/?mt=Streets&l=53481&lpw=650&cv=0&z=6&c=-43.53330%2C172.63330&e=0&al=m

Trophic levels:
Burns, N, Bryers, G, & Bowman, E (2000). Protocol for monitoring trophic levels of New Zealand lakes and reservoirs. Retrieved from
http://www.mfe.govt.nz/

Lake types:
https://en.wikipedia.org/wiki/Lake_stratification

https://en.wikipedia.org/wiki/Polymictic_lake

Phosphorus:
https://www.usgs.gov/special-topics/water-science-school/science/phosphorus-and-water#:~:text=Phosphorus%20is%20an%20essential%20element,nutrients

Nitrogen:
https://www.usgs.gov/special-topics/water-science-school/science/nitrogen-and-water#:~:text=Excess%20nitrogen%20can%20harm%20water,block%20light%20to%20deeper%20waters.